DOCUMENT RESUME

ED 129 099 FL 008 055

TITLE	Oral Language Tests for Bilingual Students: An Evaluation of Language Dominance and Proficiency Instruments.
INSTITUTION	Northwest Regional Educational Lab., Portland, Oreg.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
PUB DATE	Jul 76
CONTRACT	300-75-0329
NOTE	139p.
EDRS PRICE	MF-$0.83 HC-$7.35 Plus Postage.
DESCRIPTORS	American Indian Languages; *Bilingual Education; *Bilingualism; Chamorro; Cherokee; English (Second Language); Evaluation Criteria; Evaluation Methods; French; *Language Proficiency; Language Skills; *Language Tests; Measurement Techniques; Navaho; Oral Communication; Papago; Portuguese; Samoan; Second Language Learning; Second Languages; Spanish; Speech Skills; *Standardized Tests; *Test Reviews; Tests; Verbal Tests
IDENTIFIERS	Crow; *Language Dominance; Marshallese; Miccosukee; Yup'ik

ABSTRACT

This publication represents the first attempt to address the problem of adequate evaluation processes for testing language dominance and proficiency in bilingual education. It is produced by individuals intimately acquainted with the fields of language testing and bilingual education. Issues in language testing as well as its history are discussed. Evaluation criteria are established for tests: measurement validity, examinee appropriateness, technical excellence, and administrative usability. The following tests are described: 16 commercially available tests; 6 tests undergoing development or testing; and 2 tests used for experimental purposes. A final section provides a concise evaluation of these 24 tests according to the above-mentioned criteria. An appendix lists test development efforts in Chamorro, Cherokee, Crow, French, Marshallese, Miccosukee, Navaho, Papago, Samoan, and Yup'ik. (Author/AM)

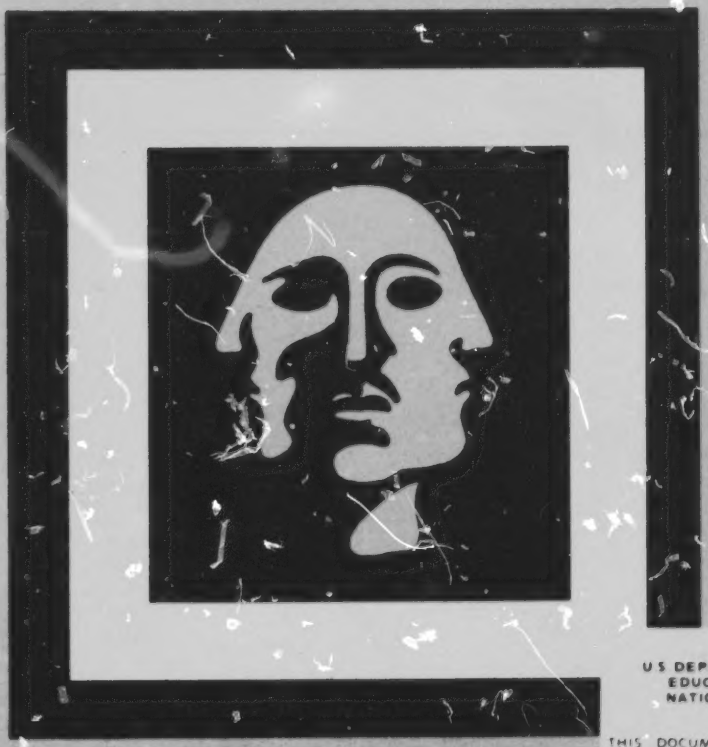
****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED129099

FL

Oral Language Tests for Bilingual Students

an Evaluation of Language Dominance and Proficiency Instruments



FL008055

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Northwest Regional Educational Laboratory

2

Oral Language Tests for Bilingual Students
an Evaluation of Language Dominance and Proficiency Instruments

Oral Language Tests for Bilingual Students
an Evaluation of Language Dominance and Proficiency Instruments

Robert J. Silverman

Joslyn K. Noa

Randall H. Russell

with Foreword by
John C. Molina

Antonio I. Fernandez, Director
Center for Bilingual Education
Northwest Regional Educational Laboratory
710 S.W. Second Avenue
Portland, Oregon 97204

Lise A. Young, editor

First printing, July 1976

This work was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare, No. 300-75-0329. The opinions expressed herein do not necessarily reflect the position or policy of the U.S. Office of Education, and no official endorsement by that office should be inferred.

Material reported herein was current as of March 1976. Subsequent editions of test instruments may be available.

ISBN 0-89354-003-X

Library of Congress Catalog Card Number: 76-21463

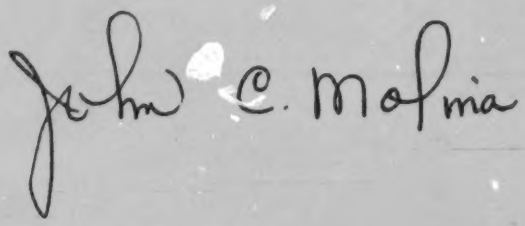
FOREWORD

One of the most rapidly growing areas within contemporary American education is the development of bilingual/bicultural curricula as a response to the needs of limited-English speaking children. In order properly to assess these needs, school districts must determine the language status of these children through some type of equitable evaluation processes. Because bilingual testing for language dominance and/or proficiency is an innovative process for most school districts, it is imperative for them to know the quantity and quality of existing and pilot tests available.

The following publication, ORAL LANGUAGE TESTS FOR BILINGUAL STUDENTS, is the first attempt to address this pressing problem for school districts. Produced by individuals intimately acquainted with the fields of language testing in particular and bilingual education in general, the publication presents concise evaluations of some twenty tests in addition to general material concerning the theory of language testing in two or more languages. Administrators, teachers, and other school

personnel involved in planning bilingual/bicultural programs in their districts will find this document an invaluable aid to identifying currently available tests and in providing objective, comprehensive evaluations of these tests in order to facilitate the selection of appropriate measurement instruments.

We sincerely hope that this publication not only will help meet present needs, but will encourage further test research, development, and evaluation in this area of innovative growth.



John C. Molina
Director, Office of Bilingual Education
Department of Health, Education, and Welfare
United States Office of Education

ACKNOWLEDGMENTS

This report reflects the contributions of many persons besides the authors and editor. We would like to acknowledge their efforts.

Antonio Fernandez, Director, Center for Bilingual Education, without whose support this document could not have been written.

J. J. Garza, Francisco Loera, and Alfredo Lugo, Migrant Education Program, Salem, Oregon, for their technical evaluations of the instruments.

Miguel Salinas, Program Director, Woodburn, Oregon, Bilingual Education Program, for his technical evaluation of the instruments.

Alfredo Aragon and Francisco Garcia, Center for Bilingual Education, for their technical evaluation of the instruments.

Tom Owen, consultant, for his help in conceptualizing parts of the document and for his technical evaluation of the instruments.

Gabriel Della-Piana, University of Utah; Leslie Palmer, American

Language Institute, Georgetown University; and Beverly Anderson, Northwest Regional Educational Laboratory, for their reviews of an earlier draft of this document.

Diane Aho, Chris Brown, Lillian Dixson, Jan Gardner, and Kathleen Peterson for their assistance in typing the preliminary documents.

TABLE OF CONTENTS

Foreword..... 5

Acknowledgments..... 7

INTRODUCTION..... 11

1. ISSUES IN LANGUAGE TESTING..... 15

   Statement of the Problem 15
   Policy of Assessment in Bilingual Education 15
   Definitions 16
   Brief History of Language Testing
   and Problems Encountered 18
   Components of Oral Language Testing 24

2. NATIONAL SEARCH FOR INSTRUMENTS..... 27

3. EVALUATIVE CRITERIA..... 29

   Measurement Validity 30
   Examinee Appropriateness 33
   Technical Excellence 37
   Administrative Usability 38

4. TEST DESCRIPTIONS..... 44

   COMMERCIALLY AVAILABLE TESTS..... 49

Language Skill Tests..... 49

Basic Inventory of Natural Language 50
Bilingual Syntax Measure 52
Comprehension of Oral Language 54
Del Rio Language Screening Test 57
Dos Amigos Verbal Language Scales 60
Gloria and David Oral Language Assessment 63
James Language Dominance Test 66
Language Dominance Criteria 69
Language Facility Test (Dailey) 71
Oral Language Dominance Measure 74
Oral Language Evaluation 76
Pictorial Test of Bilingualism and Language Dominance 78
Screening Test for Auditory Comprehension of Language 84
Test for Auditory Comprehension of Language 86

Sociolinguistic Tests..... 91

Home Bilingual Usage Estimate 92
Moreno Quick Language Assessment Inventory 94

TESTS UNDER DEVELOPMENT OR UNDERGOING FIELD TESTING..... 97

Language Assessment Battery 98
MAT-SEA-CAL Oral Proficiency Test 101
Portuguese as a Second Language, English as a Second Language Achievement Tests 106
Portuguese as a Second Language, English as a Second Language Aptitude Tests 108
Spanish/English Language Performance Screening 110
SWCEL Comparative Language Dominance Test 112

TESTS USED FOR EXPERIMENTAL PURPOSES..... 115

Ott Test of Oral Language 116
Spanish/English Dominance Assessment 118

5. TEST EVALUATIONS..... 121

Evaluation Rating Charts 123

6. SUMMARY..... 129

References..... 133

Bibliography..... 135

Appendix..... 141

INTRODUCTION

A primary goal of bilingual education programs is the utilization of two languages as media of instruction for any part or all of the school curriculum. A secondary goal is the study of the culture and history of both representative language groups. Although theoretically sound, the integration of these two goals into the school curriculum is often difficult to achieve in practice.

One critical need in attempting to realize these goals is the accurate assessment of the language preferences and relative abilities of students who might be eligible for bilingual education programs. Today there are approximately 350 federally funded bilingual education programs in the United States attempting to provide educational assistance to students who speak a primary language other than English.[1] However, only a very small percentage (3 percent) of those children eligible for some type of bilingual education program are actually receiving benefits.[2] A lack of adequate language assessment instruments has hindered the implementation of special language programs and the identification of eligible students. In

order to provide students with adequate programs and proper placement in specific classes, appropriate language measurement instruments must be identified.

As a result of the U.S. Supreme Court case of Lau v. Nichols, which unanimously ruled that Chinese American students in San Francisco had been denied a "meaningful opportunity to participate in the public educational programs,"[3] there has been increased support for bilingual education legislation at local, state, and national levels. In 1973 federally funded bilingual education projects throughout the United States numbered 209.[4] Today there are approximately 350 projects, representing an increase of almost 67 percent. Assuming that the intent of the Lau decision is to maximize equal educational opportunities for all limited-English speaking children, the development of suitable language assessment instruments to identify students actually or potentially eligible for such programs is absolutely essential.

This publication, ORAL LANGUAGE TESTS FOR BILINGUAL STUDENTS, has been produced to meet the critical need to develop and disseminate language assessment instruments and procedures. It proposes to supply school administrators and teachers, as well as directors and coordinators of bilingual education programs, with a systematic handbook of currently available language assessment procedures in order to assist them in their program planning efforts. The report contains descriptions of twenty-four tests that have been used to assess the oral language patterns of students who speak two or more languages. As part of the descriptive information, each test has been evaluated using a common set of criteria.

It is not the purpose of this report to recommend a specific test or technique for assessing oral language, but rather to provide school administrators, teachers, community members, and other people concerned with language teaching a list of currently available instruments, and descriptions of some of the important characteristics of each of these

instruments. With the information presented in this publication, individuals should be able to screen tests on the basis of their own program needs. Specific tests should then be selected for use only after thorough analysis.

The report is organized into six chapters:

1. "Issues in Language Testing": This section examines the complexities of language assessment. Although this publication as a whole focuses primarily upon oral language assessment of children, the present section discusses the entire field of language assessment, including general evaluation strategies and the strengths and weaknesses of each.

2. "National Search for Instruments": The procedures used to identify and gather the oral language tests included in this report are described in this chapter.

3. "Evaluative Criteria": This section describes the thirty-eight criteria used to evaluate the tests. The criteria were grouped into four major criterion areas: measurement validity, examinee appropriateness, technical excellence, and administrative usability. The point values assigned to each individual criterion and the points possible within every criterion area are also detailed.

4. "Test Descriptions": On the basis of their stage of development, tests were placed into one of three categories: commercially available, undergoing development and/or field testing, and experimental. This chapter provides descriptive information on each test, from administration procedures to interpretation of results.

5. "Test Evaluations": The ratings and total scores obtained by each test on the basis of the criteria are portrayed in this

chapter in matrix form.

6. "Summary": This section discusses general patterns found in the descriptions and evaluations of the instruments.

In addition to these six chapters, a supplementary section presents material not discussed in depth in the text:

"Appendix": Tests and materials that were identified as a result of the national search but which were not evaluated in this survey are listed here, together with the names and addresses of contact people involved in these development efforts.

Issues in Language Testing

STATEMENT OF THE PROBLEM

The task of accurately assessing the language and cognitive development of children who are limited speakers of English and who are in bilingual education programs in the United States is a challenging one involving several problems. Two major problems have been (1) the determination of what language skills and linguistic structures to describe and (2) the identification of adequate tools or instruments to measure the child's language usage. Bilingual education programs in the United States, since their inception in 1967, have been continuously seeking evaluation techniques for their students and programs. Unfortunately, data regarding specific language assessment procedures, teaching methodologies, and model programs are very limited.[5]

POLICY OF ASSESSMENT IN BILINGUAL EDUCATION

National policy as established in the Bilingual Education Act of 1974 states that

> . . . a national assessment [be conducted] of the educational

> needs of children who speak primarily a language other than English and of the extent to which such needs are being met from Federal, State, and local efforts, including, not later than July 1, 1976, a census of the number of children in the States and a plan to be carried out in five years for extending bilingual education programs to all such preschool and elementary school children. . .[6]

In other words, a national assessment to identify and describe specific needs of limited- and non-English speaking children has been mandated in order to carry out bilingual education programs in the United States. Supportive agencies such as the National Institute of Education have also been directed to provide research assistance in testing innovative practices, techniques, and materials in bilingual education. Assessment procedures are to serve as initial planning stages in determining both quantitative and qualitative needs in bilingual instructional programs.

Language assessment for bilingual education programs is therefore not merely the concern of isolated districts with large limited- or non-English speaking populations, but is national educational policy.

DEFINITIONS

Before describing some of the problems encountered in efforts to assess child language usage patterns, the terms language, language dominance, and language proficiency (which are used throughout this document) must be defined.

Most linguists today generally agree on the following definition of language:

> Language is a highly complex system of conventional oral symbols by means of which members of a community interact for the purpose of communication.[7]

More specifically, language is a system of three highly structured components: (1) phonology (sound), (2) grammar (form and structure), and (3)

semantics (meaning).* Oral production and aural reception are the primary factors of language; however, written systems representing the conventional meaning of the acoustic symbols in a language have also evolved. These written systems differ according to the language in question: they may be composed of letters, such as book in English, libro in Spanish, and lepblo in Chamorro; they may be represented by characters, as in Japanese; or they may be expressed pictorially:

Essential to this definition of language is the idea that language is a social tool for communication. This implies that as long as effective communication is the end result of any language, distinctions such as "standard" or "substandard" made in reference to particular dialects of a language are unimportant. Whether the language is from a child, teenager, or adult; whether it is used by individuals earning high or low incomes; whether it is spoken only by certain populations within a cultural group--language is neither good nor bad, correct nor incorrect. Systematic linguistic differences exist across all social, political, and economic levels of every culture.

Whereas language has been defined to the satisfaction of most professionals in linguistics and related fields, there is no common agreement among linguists, psycholinguists, and language test developers as to what constitutes "language dominance" and "language proficiency." Nonetheless, it is usually conceded that language dominance refers to the primary language in which a child interacts most often in the home.[8] It has also been viewed in terms of the comparison of skills in two or more languages.[9] Language proficiency generally refers to the degree to which an individual

---

*Though vocabulary (the total number of words in a particular language) is included as a separate linguistic structure within the domain of language assessment, linguistic theory generally classifies it as a subcomponent of grammar.

demonstrates his linguistic competence in a language, regardless of how that language may have been acquired.[10]

The foregoing definitions of language, language dominance, and language proficiency should be borne in mind throughout the following general discussion of language assessment.

BRIEF HISTORY OF LANGUAGE TESTING AND PROBLEMS ENCOUNTERED

Emphasis on language learning, specifically foreign language learning, emerged in the early forties within the U.S. Armed Forces.[11] Due to the military's increased awareness of the value of knowing and speaking another language brought about by exposure to foreign cultures during World War II, some twenty-seven different language courses were being conducted at fifty-five colleges and universities by 1943.[12] By 1948 the Defense Department had designed language proficiency tests in thirty-one languages for military personnel assignments.[13] Since then, other government agencies have developed tests to evaluate the language proficiency of their employees for job placement and promotion purposes.

It appears that the development, design, and in some instances the utilization of many language tests in existence today (both foreign language tests developed by the Educational Testing Service as well as those used in English as a Second Language and bilingual education programs) grew out of these earlier military and government language tests. An important characteristic of these tests used by the military and the government is that they were designed for adults and reflect the general language pattern of adults. Although adult language and child language share certain elements, there are readily identifiable differences between the two. Recent studies in Black English take cognizance of this problem.

> Once we know what we might ultimately expect in the adult speaker of the dialect, we must determine at what point

these features appear in the speech and understanding of the child learning this dialect.[14]

In assessing a child's language, the kinds of phonological, morphological, syntactic, and semantic constructs found may not be the same types of language constructs encountered in the adult. Although phonology, morphology, syntax, and semantics are general language structures frequently assessed in children, development of these domains in children may differ significantly from their development in adults. Stated another way, an adult can choose from a variety of grammatical structures to express a particular concept while a child can choose from only a few. The diagram below illustrates the differences between the degree of development of language domains in adults and in children.

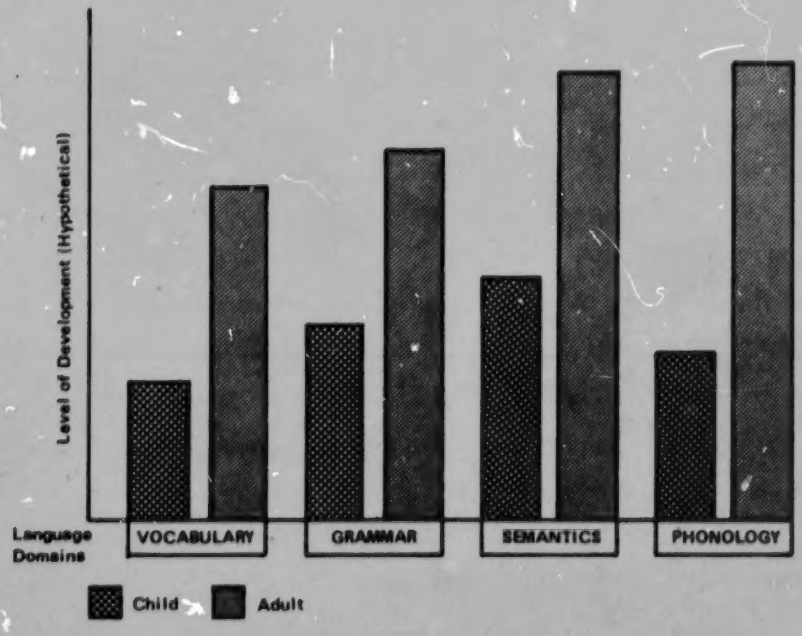


Figure 1

It is therefore essential that language assessment procedures be built on specifications of the linguistic and conceptual areas to be evaluated, and of the particular levels of language development considered for assessment. It must then be determined what evaluative techniques can be used to assess the identified areas.

The problem of what language and linguistic structures are to be evaluated in a child is difficult to resolve without first obtaining an adequate description of the linguistic rules and structures and the sociological functions of a child's language. Every child has a general cognitive set of grammatical rules and structures. The use of particular items from this set is determined by their appropriateness in specific social situations, a concept similar to what Hymes (1971) refers to as "communicative competence."[15] Information regarding a child's communicative competence should be available to test developers and test interpreters, since lack of such information may result in an inaccurate evaluation of the child's linguistic ability. For example, a child who is actually bilingual in English and Chinese might speak only English in the classroom, based on his or her perception that English is the appropriate language in this setting. Consequently, if language assessment were based solely on teacher observation in the classroom, this student might be misclassified as monolingual in English. Thus, a general descriptive pattern of the underlying communicative competence of a child, no matter what languages he or she speaks, is essential to determine what language skills and linguistic patterns to tap in evaluating a child's language.

Once a general description of the child's language usage patterns is available, the domains of linguistic structures sampled in the test must be determined. Possible domains include (1) vocabulary, (2) phonology, (3) grammar, and (4) semantics. The criteria for selecting certain domains over others may vary. In one instance, general knowledge of a large number of vocabulary words may be considered a better indicator of language skills than the ability to distinguish between parts of sound segments. The resultant

test would thus emphasize the domain of vocabulary over that of phonology. In another instance, a child who can express several abstract concepts using the same basic sentence pattern might be considered more proficient than one who is able to generate many different patterns, but is able to conceptualize only a limited number of ideas. A test designed according to this set of criteria would therefore evaluate semantic rather than syntactic elements.

Whatever linguistic structure is selected for purposes of evaluation, measurement will be influenced not only by the social context factors in a child's speech (as discussed previously), but also by selection of the communication skill area--listening, speaking, reading, writing--examined in the test. A test measuring a child's ability to comprehend oral language, for example, need not reflect the child's ability to speak that language, since these two skills are not in the same area.

The relationship among the three domains considered in language assessment (i.e., communication skills, social contexts, and linguistic structures) can be represented as follows:

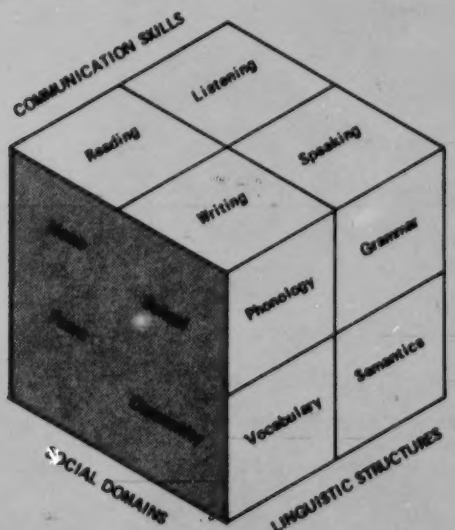


Figure 2

Once the social domains, linguistic structures, and communication skills are identified for a child or a group of children, the means to accurately measure language within these domains must be identified. Although there have been many attempts to measure the language skills of bilingual primary school children, especially in the area of speaking (Carrow, 1957; Dailey, 1968; Lambert and McNamara, 1969; Stemmler, 1967),[16] evidence suggests that there is no single best way to assess the language of children. Instead, various approaches have been utilized to measure language dominance and language proficiency in children who have limited-English abilities or who are in bilingual education programs. The major approaches used to date have been:

1. Language usage interview
2. Story telling
3. Question and answer
4. Language repetition and completion

An interview with a child and/or parent(s) can provide general sociolinguistic information about the various languages spoken and heard at home in addition to indicating with which language the child feels most comfortable. Such information can be readily verified by the child's parent or guardian. However, a language usage interview provides no information on actual language proficiency. Furthermore, social pressures to speak only one of the two languages in the home may prevent some parents from giving realistic information to interviewers. Thus, the rapport between interviewer and parent(s) will affect the accuracy of a home language usage interview.

Story telling is a simple and very popular means of assessing the general language facility of students, particularly at the elementary level. Some form of visual stimulus (e.g., a picture, photograph, painting, slide, or film) is presented to a child, who responds by telling a story about the stimulus material. This method has the advantage of enabling the child

to display a broad range of general language abilities in an atmosphere of creativity and spontaneity. However, difficulties arise in scoring each particular linguistic domain (phonology, grammar, vocabulary, or semantics) apart from the others. In general language usage all of these domains are so interdependent that separation of individual domains is very difficult. Furthermore, a specific linguistic domain may not be well represented in the child's story, and as a result, little information can be gathered regarding the child's ability within this domain. Finally, the quality of the language sample procured depends largely on the content of the visual aids used. In order to gather accurate information about the child's language ability, story stimuli should be carefully screened for linguistic and cultural appropriateness.

Another general approach in language assessment is the question-answer method which answers the need for simplicity in administration and scoring. Specific phonological, morphological, syntactical, and semantic stimuli are presented in the form of individual questions; the child must provide appropriate answers. For example, the question, "Where are the chairs?" may be presented with a picture of chairs in the classroom; the desired response would be a grammatically appropriate sentence, such as, "The chairs are in the classroom." The sequential syntactic pattern of subject, the chairs; verb, are; and prepositional phrase, in the classroom, is the response required for maximum points. (This item may also be scored as a simple vocabulary recognition test.)

Because of the specificity of stimulus questions, the question-answer approach can be used to focus on discrete parts of a particular linguistic domain. While in some cases this may be desirable, the test nevertheless measures only a small portion of overall language proficiency. In addition, a large number of test items must be included in order to ensure that each linguistic domain has been adequately covered.

Language repetition and completion affords another method of identifying specific linguistic domains of proficiency. In language repetition, a child must repeat verbal stimuli using pronunciation, intonation, and articulation equivalent to those used by the examiner. For example, the /s̬/ sound in shopping must be distinguished from the /c̬/ sound in chopping. Grammar can also be assessed using this technique by having the child complete a sentence or phrase, such as:

The girl is_____.

a) run
b) ran
c) running

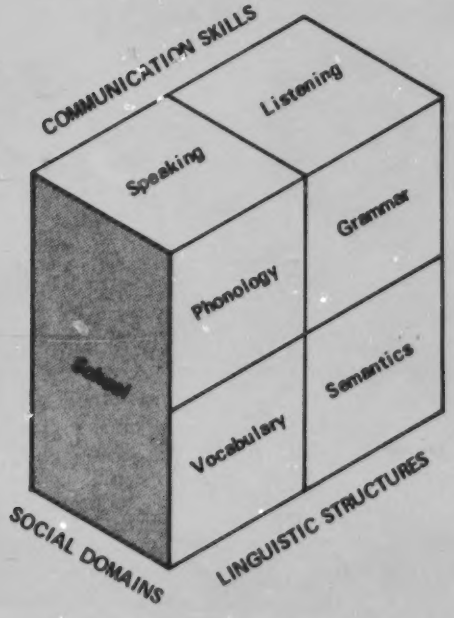
This technique provides control of responses and simplifies scoring and administration. Additionally, because specific features of the child's language are measured, information acquired can be used for diagnostic as well as placement purposes.

However, there are two major limitations to the repetition and completion method. First, a child's capacity to repeat or complete language patterns does not necessarily reflect any ability to adapt and use these patterns in everyday speech: the ability to repeat is not the same as the ability to comprehend and use a structure in another situation. Secondly, listening and speaking are so intricately related in speech that it is difficult to differentiate between the influences of each on a particular linguistic item. For instance: if a child incorrectly reproduces a phoneme, he or she either did not hear the sound(s) properly, or is unable to produce the same sound(s). It is very difficult to decide on the basis of the test alone which factor caused the mistake.

COMPONENTS OF ORAL LANGUAGE TESTING

All of the language skill tests described and evaluated in this report depict a part of the cube diagram illustrated earlier. Figure 3 displays the

portion of the language testing domain which is reflected by these instruments.



SCHOOL LANGUAGE DOMAIN

Figure 3

As illustrated in Figure 3, tests described in this report assess listening and speaking skills in two or more languages within the school environment. Tests measuring the child's performance of reading and writing skills, while they do exist, were not evaluated in this survey, since the report focuses on oral language measurement.

In the test description portion of the present survey, tests have been classified according to which of the four types of language structures they are designed to evaluate. Thus, tests may belong to one or more of

four major categories, as suggested both in Figure 3 and by the oral language domain coverage matrices included in the test descriptions:

- Vocabulary

  A vocabulary test contains a selection of one-word items which the child must identify verbally or nonverbally (for example, by pointing or marking an "x" on an answer sheet).

- Grammar

  A grammar test measures the use of morphological items (sound units which carry meaning) and the arrangement of these items together in a word and/or sentence. The child must demonstrate the ability to identify the items and/or use them in a sentence.

- Semantics

  A semantics test involves language items which may characterize conceptual domains (such as time, space, or direction relationships). A child must demonstrate the *ability to* understand concepts and/or produce meaningful sentences.

- Phonology

  A phonology test presents the distinctive sound features of the language and requires the child to recognize them, either verbally or nonverbally.

Accurate assessment of limited-English speaking and bilingual children has become national educational policy. An essential part of bilingual program planning, language assessment of any sort--language dominance, language proficiency, or sociolinguistic testing--requires a total description of the communication skills, linguistic structures, and functional usage of the child's language within all social domains. Once specific structures to be assessed have been defined, the manner to adequately evaluate these structures must be determined. It is hoped that continued research efforts in the field of child language acquisition and development will result in new, more comprehensive methods of assessing the linguistic capabilities of bilingual children.

National Search for Instruments

A major activity in the preparation of this report was the identification of currently available instruments for assessing student language dominance and language proficiency. To ensure input from all potential sources, tests were identified in three ways. First, a literature search was undertaken to identify tests, sources of tests, and current test development efforts. Second, letters requesting information about test use and development were sent to all federally funded bilingual education programs. Program directors were asked to specify what tests were being used in their programs and to list the names of any additional people or tests of which they were aware. Representatives of fifty-eight programs responded to the questionnaire. Third, telephone and personal interviews were conducted with publishers of tests, individual test developers, and other professionals involved in the assessment of bilingual students. The authors hoped that in this way additional tests and test development efforts could be identified. One result of this activity was accumulation of additional information regarding tests already collected. In a number of instances, test developers

were further requested to provide the authors with training in the administration and scoring of their instruments so that fuller understanding of the tests could be achieved.

Because the focus of this publication was the identification and evaluation of bilingual measurement instruments, not all of the approximately fifty tests collected in the national search were assessed. Of the fifty tests identified, twenty-four met the criterion of simultaneously assessing a child's oral language usage in two or more languages, and were subsequently evaluated. General kinds of language tests such as the Peabody Picture Vocabulary Test and the Pimsleur Spanish Proficiency Tests, which assess performance in only one language, were not selected for evaluation.

The criteria used to evaluate selected tests are discussed in detail in Chapter 3.

Evaluative Criteria

In order to evaluate each sociolinguistic or language skill test in a systematic and comparable manner, the decision was made to adopt a standard set of criteria that could be used for all the tests. The set of criteria included in this document is similar to that developed by the Center for the Study of Evaluation (CSE) at the University of California at Los Angeles (UCLA) in their comprehensive test evaluations. Similar criteria have been used by the Assessment Projects at the Northwest Regional Educational Laboratory (NWREL) in their evaluations of functional adult literacy tests.

Like the criteria developed by CSE and those used by the Assessment Projects at NWREL, criteria used in this report focus on four major criterion areas: measurement validity, examinee appropriateness, technical excellence, and administrative usability. Each major area consists of several individual items. Tests earned points depending upon the degree to which they met each item within the criterion areas. Points were totaled within each area and grades of good, fair, or poor were assigned

based upon total points scored in that specific criterion area.

The criteria were applied to each test evaluated in this publication. Each test was evaluated by at least two people; where differences of opinion occurred, a third person mediated the final decision. All evaluators had previous training or experience in educational measurement. Because of the various languages and cultures represented in the tests, whenever possible, each test was additionally evaluated by at least two representatives of the language and cultural group addressed by the particular test.

Information used to evaluate the tests included test manuals, other supporting documents, and other information supplied by the test authors. No attempt was made to verify available information. If necessary information either was not provided or was insufficient to permit a decision on a specific criterion item, that item scored zero points.

The remainder of this chapter details each of the criterion areas and explains bases for assigning points on each item.

MEASUREMENT VALIDITY

a. Is information provided to indicate a rigorous selection of items and careful sampling of the behavior domain?

   Information was considered adequate, if references on the construction of the test were provided. If the procedures used in developing test specifications and items were described in some detail, the test was credited with 2 points; if reference was made to the use of a specific, rigorous item selection procedure, the test was credited with 1 point; if no information was provided on item selection, the test was credited with 0 points.

b. Were any empirical procedures used for screening or selecting the items?

   Empirical procedures include item analyses, juries of experts, item difficulties, criterion-group analyses, or factor analyses.

If more than one method was reported in some detail, the test was credited with 2 points; if it was stated that one method had been used, the test was credited with 1 point; if no information was given, the test was credited with 0 points.

c. Is any information provided--either empirical or theoretical--to justify the parallelism of the alternative language forms of the test?

If item generation procedures for each language were specified indicating that the alternative language forms of the test were measuring parallel constructs, the test received 1 point. If test items were simply translated from one language to another without justification or if no item generation procedure was referenced, the test received 0 points.

d. Does the construct or type of behavior that the test purports to measure have a supportive base in linguistic, educational, psychological, or learning theory?

This criterion was applied to statements describing the theoretical basis of the test or to statements justifying the existence of the test (e.g., "Tests of grammar correlate only slightly with tests of vocabulary; therefore we felt the need for a separate grammar test, which could possibly be used as part of a more comprehensive testing effort.") If the test included such a statement, it was credited with 1 point; if not, it was credited with 0 points.

e. Has the test been employed in experiments or evaluations and produced meaningful results?

If the test scores in such experiments appeared to have yielded meaningful results, the test was credited with 1 point; if not, the test was credited with 0 points.

f. Are any concurrent validity studies reported or specifically referred to in which the criteria are related in a meaningful way to the behavior that the test assesses?

Concurrent validity studies demonstrate correlation with other criterion measures obtained at the same time as those of the test. These other criteria cannot be scores from the same test. If the criterion behavior was relevant and the coefficient was .70 or more, the test was credited with 2 points; if the coefficient was between .30 and .70, or the criterion behavior was not convincingly relevant, the test was credited with 1 point; if no study was reported, coefficients were low, or the criterion was clearly irrelevant to the nature of the test, 0 points were credited.

g. Are any predictive validity studies reported or specifically referred to in which the criteria were related in a meaningful way to the behavior which the test assesses?

Predictive validity studies provide evidence that the criterion behavior (usually successful execution of some task) may be obtained after a stated time interval. If coefficients at or above .70 were reported with relevant criteria and a time interval of one month or more, the test was credited with 2 points. If only moderate coefficients (.30 to .70) were reported, or the criteria were of questionable meaningfulness, the test was credited with 1 point. If no study was reported or referenced, or the study was irrelevant, the test was credited with 0 points.

The measurement validity ratings were summed for a total rating, varying from 0 to 11 points. These ratings were translated into letter grades according to the following scale:

9-11 points	good (G)
6-8 points	fair (F)
0-5 points	poor (P)

EXAMINEE APPROPRIATENESS

a. Does the test justify itself by explaining to the examinee in an honest manner its purpose, intent, or recommended use?

Misuse of test scores was not considered here, since such misuse is impossible to control. If the test (usually the test instructions) specifically stated the real or suggested purpose, intent, or use of the test, or if the manual suggested that such a justification be given in each situation, the test was credited with 1 point. If no purpose, intent, or use was specified; if the purpose or intent was disguised or concealed; or if examinees were led to adopt ineffective test-taking strategies, the test was credited with 0 points. This criterion was evaluated rather liberally in most cases, so that a test with instructions that began, "This is a test of your ability to spell..." was given credit for justification.

b. Are the test items personally inoffensive and appropriate in terms of difficulty for students in bilingual programs or similar settings?

If all the items appeared inoffensive and reasonably appropriate in level of difficulty, the test was credited with 2 points. If most items appeared appropriate the test was credited with 1 point. If many of the items were judged inappropriate because (1) they were ambiguous or misleading, they lacked demonstrably correct or incorrect alternatives, they were stated in unnecessarily complex language, or (2) they were too simple, or intellectually insulting in their simplicity, the test was credited with 0 points.

c. Are the items relevant, interesting and culturally appropriate for bilingual students?

This rating was made somewhat independently of test content so

that inherently interesting subject matter did not necessarily profit from this rating. One way to rephrase this rating would be to ask: Given the nature of the subject matter, have the items been made as relevant and interesting as they could be? If all the items were judged relevant, interesting, and culturally appropriate, the test was credited with 2 points. If most items were judged relevant, the test was credited with 1 point. If many of the items were judged irrelevant, dull or offensive, the test was credited with 0 points.

d. Are test instructions oral or written?

The issue here was whether successful performance on the test required only the behaviors being measured by the test items or whether competency in reading test instructions was confounded with the behaviors purportedly being measured. If instructions were either completely oral or were supposed to be read aloud in addition to being written out for examinees, the test received 1 point; if not, it received 0 points.

e. Do the test instructions describe to the examinee what he or she is expected to do on the test?

The question addressed was whether instructions precisely described all aspects of the tasks that the test measures or whether necessary issues were left unanswered or unaddressed. If all aspects were described precisely, the test was credited with 1 point. If descriptions were incomplete or left issues unanswered, the test was credited with 0 points.

f. Are the test instructions appropriate and comprehensible?

The instructions, either read by or to examinees, were inspected for appropriateness of concepts, syntax, brevity, and vocabulary in order to arrive at this judgment. If the instructions exhibited

appropriateness and comprehensibility on all counts, they were credited with 1 point; if not, they were credited with 0 points.

g. Do the test instructions provide illustrative sample items?

If the instructions included sample items that effectively clarified and accurately illustrated the task(s) involved in the test in such a way that they were truly representative of the format and difficulty of test items, the test was credited with 1 point. If there were no sample items, or if sample items presented were not representative in format or difficulty, the test was credited with 0 points.

h. Do the test pages (or materials) exhibit good layout design to facilitate perception?

Test layout was examined for effective use of perceptual organizers, such as adequate white space, regularity of item form, symmetry, clarity, and continuity. If the test page layout was clear and helpful, the test was credited with 1 point. If the layout was unclear or confusing, the test was credited with 0 points.

i. Is the physical appearance of the test of high quality?

For this rating, attention was given to the quality of the print (bold, up-to-date) and illustrations in printed tests, and the quality of sound in auditory or taped tests. If the quality was judged high, the test was credited with 1 point; if not, the test was credited with 0 points.

j. Are oral instructions or items standardized?

To meet this criterion, tests with oral instructions or oral items needed to provide either a standardized script designed to be read aloud, or a recorded version, such as a cassette tape. If the test had one or the other, it was credited with 1

point; anything less than this, such as merely suggesting topics to be mentioned to the examinee, was deemed insufficient and the test was credited with 0 points.

k. Are the time and pacing of the test appropriate?

Power tests are those that intend to measure level of performance with respect to some defined skill or ability. Tests purporting to be power tests either had to be untimed, or had to furnish evidence that 90 percent or more of the validating group attempted all items. If a test met these conditions, or was appropriately paced, it was credited with 1 point; if not, the test was credited with 0 points.

l. What is the mode of examinee response?

No points were assigned for this information, but the test or subtest being evaluated was described as requiring oral (Or), written (Wr), pointing (Po), or mixed (Mi) responses.

m. Is there a simple and direct connection between the item stem and the examinee's recording of a response?

If the mode of responding was especially simple for the examinee, such as oral response, or marking or writing directly on the test form, the test was credited with 2 points; if the test used separate answer sheets, it was credited with 1 point; if the test was complicated by the need for more than one step to get from item to answer, it was credited with 0 points.

The examinee appropriateness ratings were summed for a total rating, ranging from 0 to 15 points. These ratings were translated into letter grades according to the following scale:

12-15 points	good (G)
9-11 points	fair (F)
0-9 points	poor (P)

TECHNICAL EXCELLENCE

a. Does the test have alternative-form reliability?

The correlation between alternate forms of a test was the subject of this evaluation. If the appropriate coefficient was .90 or above, the test was credited with 3 points; if .80 to .90, the test was credited with 2 points; if .70 to .80, the test was credited with 1 point; and if below .70, or if no alternative forms were available, the test was credited with 0 points.

b. Does the test exhibit stability?

The consistency of scores as measured by test-retest reliability was the subject of this criterion. If the appropriate coefficient was .90 or more, the test was credited with 3 points; if .80 to .90, the test was credited with 2 points; if .70 to .80, the test was credited with 1 point; and if below .70, or if no stability tests were reported, the test was credited with 0 points.

c. Does the test exhibit internal consistency?

The consistency of items or parts-within-a-part as measured by split-half or Kuder-Richardson formulas was the focus of this criterion. If the appropriate coefficient was .90 or more on a sample similar to that to which the test would be used, the test was credited with 2 points; if .80 to .90, the test was credited with 1 point; and if below .80, or if no studies were reported, the test was credited with 0 points.

d. Can the testing procedures be duplicated?

A test was deemed more desirable if the procedures for administration, scoring and interpretation were sufficiently standardized so that procedures could be duplicated or replicated from the validating group. If the test provided uniformity

of procedure for administration and scoring, the gross characteristics of the standardization group were replicable, and the materials, time limits (where applicable), oral instructions, and preliminary demonstrations were precisely delineated, the test was credited with 1 point; if not, the test was credited with 0 points.

The technical excellence ratings were summed for a total rating, ranging from 0 to 9 points. These ratings were translated into letter grades according to the scale that follows:

7-9 points	good (G)
4-6 points	fair (F)
0-3 points	poor (P)

ADMINISTRATIVE USABILITY

a. Who should administer the test?

If regular program personnel, such as a teacher or aide, could read the instructions, establish rapport, and conduct the pacing, the test was credited with 1 point; if special personnel were required, such as a reading specialist, the test was credited with 0 points.

b. Is the manual clear and complete?

This criterion focused on three aspects of the test manual: discussion of the purpose, uses, and limitations of the test; clear administering and scoring directions; and description of test development and validation. If the manual's discussion, directions, and descriptions were clear and complete, the test was credited with 1 point; if not, it was credited with 0 points.

c. How many administrators or observers are needed to administer the test?

If not more than one administrator or observer was needed, the

test was credited with 1 point; if more than one was needed, the test was credited with 0 points.

d. How easy and objective is the scoring?

If the scoring was objective and simple, involving the use of a scoring guide, stencil, template, or another straightforward process such as an answer key, or if machine scoring was available, the test was credited with 2 points. If the scoring was objective but difficult, involving more than a stencil or template (e.g., scoring a passage written by the examinee for specified content), the test was credited with 1 point. If the scoring was subjective, requiring the scorer to make a non-trivial judgment, the test was credited with 0 points.

e. Who can interpret the test scores?

This rating examined whether or not regular teaching staff could interpret the test. The question was either answered by an explicit statement in the test manual or implied from the simplicity of the score conversion system. If the score could be interpreted by teaching staff, the test was credited with 1 point; if not, the test was credited with 0 points.

f. How great is the range of complexity or difficulty of the test?

Tests using some kind of grade equivalent scheme for reporting or organizing content, and having a spread of three years or more from the lowest to highest scores obtainable, or from the easiest to most difficult materials, were judged to have a reasonably extensive range, and were credited with 1 point. For tests not using grade equivalent schemes, if the validating group had a spread of three years or more on an external criterion task, or if the material in the test was organized around an extensive hierarchy (or hierarchies) of tasks, the

test was credited with 1 point. Otherwise, the test was credited with 0 points.

g. How diverse are the skills measured by the entire test?

If the test had more than one separately reported, interpretable subtest, it was credited with 1 point for diversity. If not, it was credited with 0 points.

h. How clear and simple is the process of converting the raw score into the interpreted score?

If the score conversion procedure was simple, involving one easy-to-understand step--such as a clear chart or table--or if no conversion was necessary because the raw scores were interpretable, the test was credited with 2 points. If the score conversion was complicated by lack of clear or simple tables or graphs, or if it required two or more steps to get from the raw score to the converted scores (e.g., using one table to get to another table), the test was credited with 1 point. If score conversion was necessary but complicated and (1) lacked tables or graphs, (2) required many or complicated steps (e.g., computing scores), or (3) was not explicitly provided, the test was credited with 0 points.

i. How interpretable are the scores?

This evaluation procedure looked for scores that were common, easy to interpret, and could not readily be misunderstood or misused by program personnel. If the scores were pass/fail (or some other type of binary judgment), grade equivalents, percentiles, or meaningful raw scores (e.g., a words-per-minute reading rate or a precise report of letters for which the examinee could not give the sound), the test was credited with 1 point. If the scores necessitated any ambiguous or less

common conversion, or if conversion was lacking for raw scores that were not meaningful in themselves, the test was credited with 0 points.

j. Is the validating group representative of the national population of students for whom the test was designed?

Five considerations were included in the evaluation of the composition of the validation group for the test: (1) Was the sample obtained through cluster, stratified, or random sampling rather than through incidental sampling? (2) Was the validation conducted less than five years ago? (3) Was there geographic representation? (4) Was the validating group composed of students at the appropriate educational level? (5) Were various population density characteristics (e.g., urban, suburban, rural, etc.) represented? If the answers to four or five of these questions, based upon convincing tabulation for the third, fourth, and fifth ones, was "yes," the test was credited with 1 point. If there were fewer than four "yes" answers, the test was credited with 0 points.

k. Is racial, ethnic, economic, and sex representation reported in the validation study(ies)?

If representation on more than one of these characteristics was reported, the test was credited with 2 points; if representation on only one characteristic was reported, the test was credited with 1 point; if no representation was reported, the test was credited with 0 points.

l. Are alternate forms available?

If alternate forms (developed according to the same specifications to measure the same attributes) were available, the test was credited with 1 point; if not, the test was credited with 0 points.

m. Are alternate forms comparable?

Alternate forms of instruments can be compared in many ways through considerations of content, approach or method, validities, similarity of descriptive statistics, and reliabilities. If available information indicated that the alternate forms were similar on these criteria, then the test was credited with 1 point. If a test's alternate forms exhibited low or no comparability, or the test had no alternate forms, it was credited with 0 points.

n. Can decisions be made?

This final aspect of administrative usability focused on whether the test provided information useful in making decisions concerning individual examinees. If the test manual established definite relationships between scores and specific decisions through the use of graphs, charts, cutoff scores, or other means which encouraged fairly specific decisions (e.g., "A score below this point means the examinee needs remediation to strengthen his word attack skills"), the test was credited with 2 points. If the test indicated interpretations of scores that <u>could</u> lead to specific decisions, or merely presented interpretations or definitions rather than decisions (e.g., "A high score indicates the need for testing with a standardized reading test for more accurate information"), the test was credited with 1 point. If the test provided vague or poor guidelines, leading to highly intuitive, subjective judgments, or presented no information useful in making decisions, it was credited with 0 points.

The administrative usability ratings were summed for a total rating, varying from 0 to 18 points. These ratings were translated into letter

grades according to the following scale:

15-18 points	good (G)
11-14 points	fair (F)
0-10 points	poor (P)

Test Descriptions

In compiling the tests evaluated in this survey, an effort was made to obtain as much descriptive material as possible on each instrument. The following section presents information provided about each of the twenty-four tests in summary form. Both the reviews in this section and the evaluations displayed in the matrix on pages 124-127 are organized into three categories based on their developmental status:

1. Commercially available tests
2. Tests under development or undergoing field testing
3. Tests used for experimental purposes

These classifications were made to highlight the fact that not all the tests included in this document have completed their developmental cycles. Rather than omit some tests because certain information was lacking, the authors decided to include all available oral language tests in order to identify the major test development efforts currently being undertaken in the United States. Consequently, while the same set of criteria was applied to all the tests, tests in precommercial stages generally had less technical information available than did commercially produced tests. The reader

should therefore note that the reviews and evaluations portray all instruments as they appear now. As development and validation efforts are undertaken and past studies are more fully described for precommercial tests, more complete documentation and hence more positive evaluations will result. The criterion areas of measurement validity and technical excellence are particularly likely targets for future investigation by test developers.

Within each of the three developmental categories, instruments have been further identified as either sociolinguistic tests or language skill tests. Sociolinguistic tests examine social factors that influence language. They generally provide information on the languages that are spoken in the home environment and other social settings. On the basis of data provided by these tests, a student's dominant language may be identified. In contrast, language skill tests provide information on the actual performance level of students in two or more languages. Either language dominance or language proficiency, or both, may be measured by language skill tests. When used for purposes of assessing language dominance, the tests provide information that may be used to classify students into language preference categories by comparing performance scores achieved in two or more languages. The resultant classification is based on students' relative performance in each of the languages. Language proficiency tests do not make this classification decision. Rather, they measure students' specific language skills in two or more languages, or provide a general assessment of students' overall proficiency in two or more languages. The resultant information generally is interpreted in relation to a specific content area. While comparisons between scores in two languages may be made, no language dominance classifications result from these comparisons.

In this publication, sociolinguistic tests are restricted to the section describing commercially available tests. Where tests have both a sociolinguistic and a skill component, they are described only once under one of

the two categories; however, each distinct component is evaluated individually on the matrix form.

To present the most information as concisely as possible, the summary format outlined below was adopted for all test descriptions.

TEST NAME	The test is identified. Languages evaluated by the test are enumerated directly underneath the name.
PUBLISHER	Credit is given here to the firm or individual from whom the test is available.
ORAL LANGUAGE DOMAIN COVERAGE	Displayed only for language skill tests, this matrix form identifies the aspects of oral language that the test measures. Coverage of domains is indicated by an "x" in the appropriate box(es).
DESCRIPTION	This section indicates what each instrument is intended to measure, describes any subscales included in the test, notes available alternate levels, and identifies the languages assessed by the test.
AGE/GRADE LEVEL	The age and/or grade level(s) for which the instrument is most effective are cited here.
ADMINISTRATION TIME	The total time necessary for administering the test, including initial instructions, is indicated here.
ADMINISTRATION PROCEDURES	This section specifies whether the test is administered to individuals or to groups and details the activities of both examiner and examinee(s)

during testing.

SCORING PROCEDURES	Procedures for scoring the test are described in this section.
INTERPRETATION PROCEDURES	Interpretable results obtained from administering the test are described and types of conclusions that can be drawn (or decisions that can be made) on the basis of test results are specified.
VALIDITY	Information supporting the validity of the test appears in this section.
RELIABILITY	Evidence on behalf of the reliability of the test is presented.
FIELD TRYOUTS	In instances where information is available on the nature of field tryouts conducted with the tests, it has been included in this section. Characteristics of the tryout population(s) are included if test developers reported such material.
COST	Prices for commercially available instruments are quoted here.
RATINGS	This section notes the page numbers on which evaluations of the instrument appear.

Commercially Available Tests:
Language Skill Tests

BASIC INVENTORY OF NATURAL LANGUAGE
English/Spanish

PUBLISHER

Chess and Associates, Inc.
Educational Systems and Services
P. O. Box 833
San Bernardino, California 92401
(714) 883-3093

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening				
Speaking		X		

DESCRIPTION

The Basic Inventory of Natural Language (BINL) is a criterion referenced system designed to assess the language dominance, proficiency, and growth of students in school language arts and reading programs. Scales identified within the test include: (1) an index of language ability, (2) average sentence length, (3) average level of complexity, and (4) fluency (total number of words). The languages dealt with in the manual are Spanish and English. From individuals' scores, a Class Oral Language Profile can be computed.

AGE/GRADE LEVEL

Age levels not specified; however, the BINL has been used with grades K-2

AVAILABILITY OF ALTERNATE FORMS

No alternate forms are available.

ADMINISTRATION TIME

Not specified, but approximately two weeks training required

ADMINISTRATION PROCEDURES

The BINL is designed to be administered in small groups. The aim of administration is to elicit a natural language sample; therefore, it is essential that the atmosphere not be one typical of a testing situation. The three activities suggested as

means of obtaining natural language samples must be incorporated into the curriculum prior to taking language samples so that children are familiar with the procedures used. Language samples are tape-recorded. Children are asked to make up a story using either a story starter poster, story sequence pictures, or talk tiles (provided in the test kit). Each in turn tells his/her story to the group. The examiner should be as unobtrusive as possible. Other children in the group should be encouraged to make comments or ask questions concerning the stories. Ten sentences (partial and/or complete) from each student comprise a minimum sample. For bilingual students, a sample should be taken in each language.

MATERIALS USED

examiner: tape recorder, BINL instruction manual, Individual Oral Scores (IOS) sheets, Individual Oral Language Profile (IOLP) sheets, Class Oral Language Profile (COLP) sheets

examinee: story starter posters, talk tiles, or story sequence pictures

SCORING PROCEDURES

Student samples are transcribed onto the IOS sheets. Use of nonstandard or dialectal forms does not affect a student's score. Complete or partial sentences, words, modifiers, phrases, and clauses are tabulated for scoring purposes. Each type of item is assigned a point value. At the conclusion of the test, word total, sentence total, average sentence length, and average level of complexity are computed for each child. Scoring is based on simple addition and division of scores derived previously. IOLP sheets can be filled out on the basis of this information; a COLP sheet can be filled out from class IOLPs. Data are merely transferred from one type of sheet to another.

INTERPRETATION PROCEDURES

The developer states that by comparing a student's IOLPs in both languages his or her dominant language can be identified.

VALIDITY	No studies are reported.
RELIABILITY	No studies are reported.
FIELD TRYOUTS	Normative data on the BINL are being assembled. A preliminary study is reported in which the BINL was administered to 346 children, from kindergarten to second grade, who spoke both English and Spanish.
COST	Kit - $85.00
RATINGS	See pages 124-125.

BILINGUAL SYNTAX MEASURE
English/Spanish

PUBLISHER	Harcourt Brace Jovanovich, Inc. California Office Test Department Polk and Geary Streets San Francisco, California 94109 (415) 771-3600

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening				
Speaking		X		

DESCRIPTION	The Bilingual Syntax Measure (BSM) is designed to measure children's oral proficiency in English and/or Spanish grammatical structures by using natural speech as a basis for making judgments. It yields information on language dominance and structural proficiency in English and/or Spanish.
AGE/GRADE LEVEL	Ages 4-9; grades K-2
AVAILABILITY OF ALTERNATE FORMS	No alternate forms are available.

ADMINISTRATION TIME	Approximately 10-15 minutes
ADMINISTRATION PROCEDURES	The BSM is administered to individual children. The examiner must be able to speak and write whichever language is being tested. The examiner should put the child at ease, ask the preliminary questions, and then proceed with the body of the test. During a test session, the examiner shows pictures, asks questions, and records the student's responses on the scoring form. The student responds orally to all questions.
MATERIALS USED	examiner: picture booklet, "Child Response Booklet" examinee: none
SCORING PROCEDURES	Responses are awarded one point if grammatically correct and zero points if grammatically incorrect, or if the child either does not respond or gives a response of "I don't know." Pronunciation is not scored and is disregarded for the purposes of this test. A guide of what is grammatical for the BSM is included in the English and Spanish manuals
INTERPRETATION PROCEDURES	Children are classified into one of five proficiency levels on the basis of scores obtained on the BSM. Procedures for classifying a child, described in the "Child Response Booklet," are objective and straightforward. The levels for the English version are: (1) no English, (2) receptive English only, (3) survival English, (4) intermediate English, and (5) proficient English. Spanish levels are comparable. Precise descriptions of each level are provided in the test manual. A comparison of proficiency levels in two languages yields information that indicates the dominant language and degree of bilingual proficiency. Instructional suggestions are offered for children scoring at each level.
VALIDITY	The test developers believe that of all the linguistic domains, syntax varies least across

speakers of a particular language; thus, syntax was chosen as the vehicle for item selection and scoring on the BSM. Items were chosen in such a manner that score is not affected by general information or cultural background.

RELIABILITY — No studies are reported.

FIELD TRYOUTS — The BSM has been administered to 1,956 pupils: 663 in kindergarten, 663 in grade 1, and 630 in grade 2. Approximately 70 percent of the pupils were tested in both languages (1,139). The ethnic breakdown of the 1,572 pupils for whom information was available was:

Mexican	749
Anglo American	375
Puerto Rican	169
Cuban	133
Afro American	80
Other Spanish	45
Other	21
	1,572

COST — Kit - $45.00

RATINGS — See pages 124-125.

COMPREHENSION OF ORAL LANGUAGE - LEVEL I
English/Spanish
Inter-American Series

PUBLISHER — Guidance Testing Associates
6516 Shirley Avenue
Austin, Texas 78752
(512) 452-6969

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening	X	X		
Speaking				

DESCRIPTION	The Comprehension of Oral Language - Level I (COL-1) test is designed to assist in the appraisal of language ability in the early stages of formal instruction. It provides an estimate of a child's ability to understand short verbal expressions presented orally. Two versions of comparable measurement are available in both Spanish and English (CO-1-LES, CO-1-MES; and OC-1-LE, OC-1-ME, respectively).
AGE/GRADE LEVEL	Ages not specified; grades K-1
AVAILABILITY OF FORMS	One alternate form is available for both the Spanish and the English versions of the test.
ADMINISTRATION TIME	Not specified; estimate approximately 20 minutes
ADMINISTRATION PROCEDURES	The COL-1 can be administered to groups of from eight to twelve children. It may also be administered to individual children who are unable to function in a group testing situation. The examiner must be bilingual to administer both forms of the test. The examiner reads questions to the examinees, who respond by marking the appropriate picture in their test booklets.
MATERIALS USED	examiner: administrator's manual examinee: test booklet (pictures), pencil
SCORING PROCEDURES	Examinee responses are compared to the correct answer for each question contained in the administrator's manual. One point is given for each correct response. Practice questions do not count. The total number of correct responses is recorded on the front of the test booklet.
INTERPRETATION PROCEDURES	Comparison of a child's scores on the English and Spanish versions provides some insight into relative abilities in each language. The designers of the test strongly recommend the construction of regional or local norms by those who use the test. The COL-1 is a recent addition to the Inter-American Series and the statistical information available now for use in interpreting

	scores is limited. Contact the publisher for additional developments
VALIDITY	Content validity: Exhaustive procedures have been used in developing content for the Inter-American Series. These include item analysis, juries of experts, item difficulties, and item frequencies. The content development of the COL-1 is not referenced; however, many items are taken from tests of general ability and reading in the Inter-American Series, which have been referenced.
RELIABILITY	Alternate form scores have been compared in scatter diagrams in the Report on New Tests of Oral Comprehension; however, correlation coefficients have not been computed.
FIELD TRYOUTS	Information supplied on COL-1 tryouts consists of data reported on charts and graphs. One tryout group was composed of fifty-five kindergarten children, of whom twenty-six were Spanish speaking and twenty-nine English speaking. The OC-1-LE and ME versions were administered to a group of fifty-two first grade children from four school districts who were either Spanish speaking or Spanish surnamed. The CO-1-LEs and MEs were administered to a group of seventy-eight first grade children from six school districts who were either Spanish speaking or Spanish surnamed. All four versions were administered to a group of sixty-three Spanish surnamed children. Results of these tryouts are presented in scatter diagrams.
COST	Test booklets (ten or more) - $.08 each
RATINGS	See pages 124-125.

DEL RIO LANGUAGE SCREENING TEST
English/Spanish

PUBLISHER

National Educational Laboratory Publishers, Inc.
P. O. Box 1003
Austin, Texas 78767
(512) 385-3588

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening	X	X	X	
Speaking		X	X	

DESCRIPTION

The Del Rio Language Screening Test (DRLST) was devised to identify children with deviant language performance in English and Spanish who need further evaluation. In addition, it may be used as a diagnostic instrument and as a test of relative bilingual proficiency. It is designed to group children into three language categories: norms have been established for each of these three groups: (1) English speaking Anglo American children, (2) predominantly English speaking Mexican American children, and (3) predominantly Spanish speaking Mexican American children.

The DRLST is comprised of five separately reported subtests. The "Receptive Vocabulary" subtest assesses comprehension of single nouns and verbs. The "Sentence Repetition Length" subtest measures memory for a string of words related in a grammatical sentence. The "Sentence Repetition-Complexity" subtest assesses ability to repeat relatively short sentences which gradually increase in grammatical complexity. The "Oral Commands" subtest evaluates memory for increasing numbers of oral commands. The "Story Comprehension" subtest assesses a child's memory for information presented in brief stories.

AGE/GRADE LEVEL

Ages 3 to 7

AVAILABILITY OF ALTERNATE FORMS	No alternate forms are available.
ADMINISTRATION TIME	Not specified
ADMINISTRATION PROCEDURES	The test is administered to individual students. To administer both parts of the test, the examiner must be bilingual since all instructions are given orally. The manual states that every response should be reinforced by an encouraging remark. The examinee responds in one of three ways, depending on which section of the test is being administered: (1) by pointing to a picture (in the test manual), (2) orally, or (3) by performing specified actions requested by the examiner (oral commands).
MATERIALS USED	examiner: test manual, scoring form, pencil, book, paper, chair, table examinee: pencil, book, table, chair, paper
SCORING PROCEDURES	Correct responses to the various tasks in the five sections of the DRLST are specified in the test manual. The examiner marks the scoring form, giving one point for a correct response and zero points for an incorrect response.
INTERPRETATION PROCEDURES	The purpose of the DRLST is to identify children with deviant language performance who need further evaluation. By comparing the child's score to the norms provided in the appropriate table, his or her scores can be interpreted. A cutoff score is specified (at or below the second standard deviation below the mean) for deviant performance. Scoring patterns on specific items can also be used to locate problem areas for the child. In addition, the manual presents guidelines for determining bilingual proficiency of Mexican American students through administration of both forms of the test and special interpretations of the norm tables.
VALIDITY	Content validity: Scores on the DRLST for each of the three main groups were compared. It was

established that the norms for each of the three language categories differed significantly (F ratios calculated and significant beyond the .001 level of confidence). Thus, the three main group norms cannot be compressed into a single norm.

Developers of the test believe that the five tasks included in the test assess some of the basic parameters of language development. Items included in the test were chosen after consideration of the language and cultural background of the Del Rio area. Bilingual educators in the area were consulted. The "Sentence Repetition-Complexity" test was organized using syntactic hierarchies established in cited research studies. The "Story Comprehension" subtest was field tested. Final selections were based on how well items discriminated between age groups.

RELIABILITY

Test/retest: From each main group thirty-two subjects were randomly selected for retesting exactly two weeks after the initial administration of the test. Reliability was computed using Spearman's rank-order correlation procedure for matched groups. The correlation coefficients for each main group and each subtest are presented in the manual. All coefficients are significant beyond the .01 level of confidence when converted into t values.

FIELD TRYOUTS

The test was standardized on the three main groups mentioned in the description. The standardization group consisted of 384 children between the ages of three years and six years eleven months. One hundred-and-twenty-eight normal children were selected for each group and were equally distributed across the age range. All subjects were randomly selected from Del Rio, Texas, and roughly represented the distribution of ethnic groups in that town. Subjects met four criteria: (1) All had normal hearing; (2) Case histories of all children showed no causes of learning difficulties; (3) All children were informally assessed as normal; and (4) Language group placement was made only

if subjects spoke that language more than 75 percent of the time at home.

COST — Copy - $6.00

RATINGS — See pages 124-125.

DOS AMIGOS VERBAL LANGUAGE SCALES
English/Spanish

PUBLISHER — Academic Therapy Publications
1539 Fourth Street
San Rafael, California 94901
(415) 456-1394

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening	X			
Speaking	X			

DESCRIPTION — The Dos Amigos Verbal Language Scales (DAVLS) identifies language development at the comprehension level in both English and Spanish by using vocabulary items. It is designed for use with Spanish/English bilingual-bicultural children. Two scores are yielded, Spanish and English, which are interpreted both separately and in conjunction with each other.

AGE/GRADE LEVEL — Ages 5 to 12 years 6 months

AVAILABILITY OF ALTERNATE FORMS — No alternate forms are available.

ADMINISTRATION TIME — Approximately 10 minutes for each section; 20 minutes total

ADMINISTRATION PROCEDURES — The test is administered to individual children by a bilingual examiner. The examiner explains the task requirements (i.e., giving opposites for the stimulus words); the examinee responds orally.

MATERIALS USED	examiner: test manual, scoring sheet examinee: none
SCORING PROCEDURES	The examiner scores responses on the examination sheet. Correct responses, as described on the sheet, receive a plus, all others a zero. The number of correct responses on each scale (English or Spanish) are totaled and recorded on the front of the examination sheet as raw scores.
INTERPRETATION PROCEDURES	Tables are provided to allow interpretation of raw scores. Two tables give the means and standard deviations for the standardization group by grade level and age range. The appendix in the manual consists of age tables for English and Spanish which give percentile rankings for raw scores. The percentile indicates the child's language development as compared to a random sample of children of similar background and age. One of two determinations may be made on the basis of a child's scores: needs intensive oral language instruction, or needs further evaluation (conceptual weakness). Procedures for making these determinations are provided in the test manual.
VALIDITY	Content validity: In the theory behind the construction of the DAVLS, it is stated that "in order to obtain a measure of the bilingual child's comparative English/Spanish language development, the instrument employed must elicit meaningful responses at the comprehension or conceptual level in both languages." To measure development at this level, paired responses were chosen. A word and its opposite comprise each test item. Particular items were chosen from oral language activities in kindergarten through sixth grade. Items were ranked according to difficulty and field tested; compensations were made for alternative responses. A panel of consultants in Mexican American culture reviewed test items.

Concurrent validity: Thirty children referred for reading failure were tested on the DAVLS, the Marianne Frostig Developmental Test of Visual Perception, the Gilmore Oral Reading Test, an informal listening test and the Keystone Telebinocular Visual Survey. It was stated that the DAVLS results compared well with the other tests, although no correlations were reported in the manual. The DAVLS was also favorably compared with the conceptual abilities score derived from the Wechsler Intelligence Scale for Children, although again no correlation studies were reported.

RELIABILITY

No studies are reported.

FIELD TRYOUTS

The DAVLS was field tested with more than three hundred children in three clinical situations. The manual directs those interested in acquiring further information on field test procedures to address inquiries to the author.

Children were selected for field testing only if they met three criteria: (1) They were of Mexican American ancestry; (2) Spanish was the first language learned; and (3) They had been reared in the (non-English) cultural milieu. Examiners were bilingual and met the same criteria as the children. A total of 1,224 children from kindergarten through sixth grade made up the standardization population. All were residents of the south Texas area.

COST

Kit - $6.50

RATINGS

See pages 124-125.

GLORIA AND DAVID ORAL LANGUAGE ASSESSMENT
English/Spanish

PUBLISHER

Institute for Cultural Pluralism
School of Education
San Diego State University
San Diego, California 92182
(714) 286-5000

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening		X		X
Speaking		X		X

DESCRIPTION

This instrument was used by the Bilingual Education Program in Pueblo, Colorado, to provide a diagnostic analysis for individual students. The analysis includes grammar and phonetics. The assessment is a sentence repetition task. Two versions of the instrument are available, Spanish/English and English. The instrument can be used to evaluate Black language as well as Spanish speaking groups.

AGE/GRADE LEVEL

Grades K-2; publisher states that test is also effective at the junior high school level

AVAILABILITY OF ALTERNATE FORMS

No alternate forms are available.

ADMINISTRATION TIME

Approximately 9 minutes

ADMINISTRATION PROCEDURES

The test is administered to individual children. A specially designed machine (Sentinel 16 by LaBelle) plays a tape recording of a sentence, which is heard by the student on a headset, while simultaneously showing a related picture from a filmstrip cartridge. The student repeats the sentence he or she has just heard into a microphone. These responses are recorded on tape with a time lapse between each one.

MATERIALS USED	examiner: Sentinel 16 testing machine, filmstrip cartridge and tape of test sentences, class summary forms for Oral Language Assessment Profiles, checksheet of test sentences examinee: microphone, tape recorder
SCORING PROCEDURES	Special training as specified in the manual is necessary to score the instrument. This use of the Gloria and David Oral Language Assessment test was developed in a program designed to train teachers to perform meaningful language evaluation. The program includes training in the use of the Sentinel 16, administration of the Gloria and David test, and special training in dialects, phonology, and grammar for English, Spanish, and Black language. Numerical scores are not derived; rather, an individualized language profile is provided for each student.
INTERPRETATION PROCEDURES	From each language profile, an analysis of performance can be made that indicates grammatical and phonological problem areas for the student. It is also stated that the profile indicates language dominance and comprehension.
VALIDITY	Content validity: A content analysis of the two tests was performed. Enough items were found both phonological and grammatical to allow evaluation of Spanish and English oral language within certain limitations.
RELIABILITY	In a research study the instrument was checked to determine if ratings derived from it would be consistent. Children in grades K-2 in San Antonio, Texas, were tested on the Gloria and David. From a group of 750 recordings, those of ten Black and ten Mexican American children were chosen as representative samples. The Black language samples were evaluated by fifteen experts; the Mexican American samples were evaluated by fourteen experts. By assigning numbers to the experts' scaled ratings for specified areas of language assessment (ten areas for Black language, fifteen areas for

Spanish), it was possible to compute the mathematical reliability of the experts' judgments. The index ranged from 0.0 (no reliability) to 1.0 (perfect reliability); anything lower than .80 was considered to be of questionable reliability. Reported results show high to moderate reliabilities within the instrument for all but three areas. In addition to good agreement on ratings, it was found that the bases for evaluation made by the experts were consistent.

Examination of the mean ratings assigned on each scale to each child was reported to show that the ratings were highly differentiated within each child's performance and between children (no figures quoted).

FIELD TRYOUTS

The Gloria and David Oral Language Assessment instrument was used in a bilingual education program in Pueblo, Colorado. A total of 670 children in the program, grades K-1, and sixty-two children in Head Start and preschool programs was tested and evaluated.

Teachers received individual student profiles and charts that indicated areas of difficulty for each child. No validity or reliability studies are reported for this test group.

A random sample of one hundred kindergarten pupils' responses was submitted for further analysis. The tapes of their performance were evaluated using the scale which had been used in a previous study (see RELIABILITY). Spanish and English usages were rated on dominance, comprehension, production, phonology, and syntax; English was also judged on inflection.

COST

Cost not available

RATINGS

See pages 124-125.

JAMES LANGUAGE DOMINANCE TEST
English/Spanish

PUBLISHER

Learning Concepts
2501 North Lamar Boulevard
Austin, Texas 78705
(512) 474-6911

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening	X			
Speaking	X			

DESCRIPTION

The James Language Dominance Test is an English/Spanish test designed to determine language dominance. Five categories of dominance are identified: (1) Spanish dominant, (2) bilingual with Spanish as home language, (3) bilingual with English and Spanish as home languages, (4) English dominant, but bilingual in comprehension, and (5) English dominant. This test measures both comprehension and word production in Spanish and English.

AGE/GRADE LEVEL

Grades K-1

AVAILABILITY OF ALTERNATE FORMS

No alternate forms are available.

ADMINISTRATION TIME

7-10 minutes

ADMINISTRATION PROCEDURES

The test is administered to individual children. All examiners must be fluent in both Spanish and English. The examiner asks questions first in Spanish, then in English; the examinee responds both orally and by pointing to pictures in the test book. The examiner notes incorrect responses on the scoring form.

MATERIALS USED

examiner: test manual, test book, scoring booklet
examinee: none

SCORING PROCEDURES	Comprehension: The examiner circles incorrect responses on the scoring form. Correct responses for each language are counted and entered separately on the scoring form to determine total comprehension score. Provisions are made for possible dialectal variations. Production: Incorrect responses are circled on the scoring form. Alternate responses are recorded on the form and later checked against an alternate response chart. Provisions are made for possible dialectal or phonological variations. The production scores for English and Spanish are then recorded on the scoring form.
INTERPRETATION PROCEDURES	The test manual provides a language classification guide. Using the examinee's scores and the formulas provided, one can determine which of the five specified categories of language dominance a student may be placed in.
VALIDITY	Content validity: Items on the test were selected in two ways: (1) Successive versions of the test were developed. Items unfamiliar to the students or subject to dialectal variation were eliminated. (2) An item pool was developed by the author while working intensively with one class. Items were gathered from workbooks or the child's home environment. Predictive validity: The first test version was administered to three first grade classrooms. These classes were then grouped according to the results of the test. After two months, a spot check revealed that all teachers felt the children had been appropriately placed. Construct validity: This was approached in three ways. 1. Group norms: A table was made showing the distribution of means and standard deviations for a sample of test

children grouped in the five language categories. There was considerable overlap of adjoining groups, but good separation of more distant groups.

2. Expectancy tables: These five tables were designed to show the specific distribution of percentages of children tested around particular cutoff scores used to compute the grouping formula.

3. Multiple discriminant analysis: A multiple discriminant analysis was conducted on the five scales defined by the test. It was found that the groups are empirically clustered in such a way that the probabilities of misclassification of a particular student were less than .1 in all but one case.

RELIABILITY

The first version of this test was administered to three first grade classrooms in Austin, Texas. The children were grouped according to test results. Six months later these same children were tested on a second version of the test. The classifications of the second version were in 98 percent agreement with the classifications of the first version.

FIELD TRYOUTS

Testing with the first and second versions has been described. In both instances, the test was administered in Austin, Texas. The children were from a low income neighborhood and from varied language backgrounds, ranging from Spanish dominant to English dominant. A majority were English dominant.

The third version of the test, from which the construct validity evidence is derived, was administered to 464 Mexican American children in Del Valle and Weslaco, Texas, in the spring of 1974. The children in Del Valle were from backgrounds similar to those in Austin. The majority of the children in Weslaco, a town in the lower Rio Grande Valley, came from Spanish speaking environments.

COST — Kit - $16.95

RATINGS — See pages 124-125.

LANGUAGE DOMINANCE CRITERIA
English/Spanish

PUBLISHER — Fernando Canedo, David Gustafson, Americo Lopez-Rodriguez
Bilingual-Bicultural Education Program
California State University
Fullerton, California 92634
(714) 870-3994

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening				
Speaking		X	X	

DESCRIPTION — This instrument is designed to determine which language (English or Spanish) is dominant in the child's environment and to provide a rapid measurement of the child's linguistic abilities in these languages. There are two parts to the test: a social section and a performance section. Four scales are derived from test results: (1) comparative language pressures, (2) comparative language strengths, (3) Spanish performance score, and (4) English performance score.

AGE/GRADE LEVEL — Grades K-1

AVAILABILITY OF ALTERNATE FORMS — No alternate forms are available.

ADMINISTRATION TIME — Not specified; estimate 10-15 minutes

ADMINISTRATION PROCEDURES — The test is administered to individual children. The examiner must be bilingual. The first section of the test consists of a series of ques-

	tions, asked and answered orally, concerning the language the child uses with his family. The examiner notes responses on a scoring sheet. The second section of the test evaluates language production. It is administered first in Spanish, then in English. The examinee responds orally. The examiner rates responses on a scoring sheet.
MATERIALS USED	examiner: test section, answer sheets 1 and 2 examinee: none
SCORING PROCEDURES	Social dominance scale: Responses to questions on language usage are assigned points according to which language is used. Scores on the Spanish and English sections are subtotaled. The English total is subtracted from the Spanish total. This figure is then entered on the scoring sheet and the result is noted as positive or negative. Performance scale: Responses to questions are scored according to criteria on the scoring form. Scores on the Spanish and English sections are subtotaled. The English score is subtracted from the Spanish and the value of the resultant figure, whether positive or negative, is noted.
INTERPRETATION PROCEDURES	Social dominance scale: Using the chart provided, the number derived from the scoring procedures is interpreted to determine comparative language pressures. One of three ratings results: (1) monolingual Spanish, (2) bilingual, or (3) monolingual English. Performance scale: In interpreting these scores, three charts are used. The number derived in the scoring procedures is used to determine comparative language strength. The subscores, Spanish performance and English performance, are compared to their respective charts to determine proficiency in each language.

VALIDITY — No studies are reported.

RELIABILITY — No studies are reported.

FIELD TRYOUTS — No studies are reported.

COST — Copy - $1.00

RATINGS — See pages 124-125.

LANGUAGE FACILITY TEST
English/Spanish

PUBLISHER — The Allington Corporation
801 North Pitt Street #701
Alexandria, Virginia 22314
(703) 683-3187

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening				
Speaking		X	X	

DESCRIPTION — The Language Facility Test (LFT) was designed to measure ability to use oral language. There are two different scoring systems. The first yields a measure of how well the individual is able to conceptualize and communicate in his or her chosen language. The second scoring system gives a diagnostic profile of ability to speak standard English. The test is designed for use with individuals who speak English or a dialect of English, but has been used in bilingual settings.

AGE/GRADE LEVEL — Norms provided for ages 3 to 15

AVAILABILITY OF ALTERNATE FORMS — Three pictures are used to administer the LFT. There are three alternate groups of pictures provided.

ADMINISTRATION TIME	Approximately 10 minutes
ADMINISTRATION PROCEDURES	The LFT is administered to individual students. The examiner reads a set of instructions to the examinee and then shows one of a series of three pictures to him or her. The examinee must then tell a story about each picture. The examiner may prompt, but only according to the specified script. This procedure is repeated for the next two pictures. The examiner records all responses.
MATERIALS USED	examiner: test manual, three test pictures, tape recorder examinee: none
SCORING PROCEDURES	1. Responses to each of the three pictures are scored on a scale from zero to nine. The maximum possible score for the three pictures is twenty-seven. The scale is independent of standard English vocabulary, enunciation, information, or grammatical exactness. Language structure is evaluated, however. Some factors used in scoring are story completeness, organization, verbs of action, and use of past or future tenses. Scoring instructions, criteria, and examples are provided in the test manual. 2. A twenty-four-category code of common deviations from standard English was devised. Clerks, trained to recognize the specified deviations, listened to the taped stories and tallied the deviations as they occurred.
INTERPRETATION PROCEDURES	1. Total raw score for the three pictures can be interpreted by using any or all of the three charts provided. One chart yields information on the language facility development of the student for his or her age (accelerated, high-average, low-average, slow). Another compares a student's language facility age equivalent with chronological age. The last chart shows

percentile norms based on 4,000 cases from ages three to twenty (the actual chart includes percentile norms only for ages three to fifteen). From this chart the percentile equivalent for a student's score in his or her age group can be read.

2. A diagnostic profile can be obtained by computing the frequency of each of twenty-four common deviations from standard English.

VALIDITY

Content validity: Pictures were chosen for inclusion if they demonstrated in tryouts that they could be meaningful to students from a great variety of backgrounds. Statistical analyses have been made which demonstrate that the pictures give similar results and measure the same thing. Data regarding these analyses are not included in the manual.

Concurrent validity: The LFT has been compared to a variety of other instruments (e.g., the Stanford-Binet Intelligence Scale and the Metropolitan Reading Readiness Test). Also, a teacher rating of scholastic performance was used. The correlation coefficients ranged from -.01 to .28.

RELIABILITY

Test/retest: A number of studies have been conducted with various groups and over a variety of time spans. Interim activities were of an educational or remedial nature. Groups tested include trainable mental retardates, deaf children, nursery school students, prekindergarten children, and third graders. The correlation coefficients ranged from .46 to .9[illegible].

The LFT has also been checked for inter-rater reliability. In one study 116 tests were scored independently by two different scorers. The Pearson product-moment correlation between the two was .917. Other samples of one hundred or more have yielded scores ranging from .88 to .94.

FIELD TRYOUTS

Several studies have been conducted using the LFT. A list of these studies is provided in the manual, and transcribed protocols, including such information as age, sex, and cultural status, are available from the author for analysis.

Groups assessed by the instrument have included: (1) mental retardates, ages 4-16; (2) physically handicapped, ages 6-16; (3) poor readers, ages 12-15; (4) Spanish American bilinguals (inner city or small town); (5) inner city preschool children; (6) inner city students in grades 1-12; (7) deaf children, ages 6-18; (8) middle-class, nursery school, and first grade students; (9) children with behavior problems, ages 12-15; (10) inner city custodial children, ages 6-12; and (11) professional adults.

COST

Kit - $15.00

RATINGS

See pages 124-125.

ORAL LANGUAGE DOMINANCE MEASURE (PAL)
English/Spanish

PUBLISHER

Primary Acquisition of Languages: A Dual Language Program
The El Paso Public Schools
Department of Curriculum and Staff Development
El Paso, Texas 79998
(915) 779-4152

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening				
Speaking		X		

DESCRIPTION

This test is designed to measure Spanish/English language dominance. There are two language versions, English and Spanish, which yield separate scores. The test is

	scored on language structure and yields some inferential diagnostic information.
AGE/GRADE LEVEL	Primary grades; field tested with grades K-3
AVAILABILITY OF ALTERNATE FORMS	No alternate forms are available.
ADMINISTRATION TIME	Approximately 14 minutes total for both parts
ADMINISTRATION PROCEDURES	The test is given to individual children by a bilingual examiner. The examiner is instructed to put the child at ease and to converse with the child during the test. Test questions are administered orally and answered orally. The examiner notes the responses of the examinee in the test booklet. Complete instructions for examiner questioning procedures are provided.
MATERIALS USED	examiner: examiner's manual, picture booklet, scoring booklet (English and Spanish) examinee: none
SCORING PROCEDURES	Complete guidelines are provided for recording responses. Scores are computed by assigning a numerical value to each response (procedures for scoring language structure are included in the manual). Scores are recorded in the test booklets.
INTERPRETATION PROCEDURES	Raw test scores are interpreted by two tables. The first assigns a numerical level from one to five for each of the language components of the test. The second table uses this information to determine language dominance. This information can then be used as a placement tool. The test also provides a limited amount of diagnostic information on ability to handle language structure.
VALIDITY	No studies are reported.
RELIABILITY	The test has been used for three years in grades K-3 in the El Paso Public School System and is

claimed to have yielded reliable results.

FIELD TRYOUTS — This test has been used for three years in grades K-3 on approximately 13,069 children. Results are described as reliable.

COST — Set - $3.00

RATINGS — See pages 124-125.

ORAL LANGUAGE EVALUATION
English/Spanish

PUBLISHER — D. A. Lewis Associates, Inc.
7801 Old Branch Avenue
Suite 201
Clinton, Maryland 20735
(301) 868-7070

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening				
Speaking		X		

DESCRIPTION — The Oral Language Evaluation (OLE) test consists of two parts: Part I is primarily a screening device to identify children who need training in a second language; Part II is designed to evaluate the individual's production of oral language structures. Part II has eight subscales that register production of specific language structures (i.e., subject-verb-complement sentences). This test is designed for use in English and Spanish.

AGE/GRADE LEVEL — Ages not specified; apparently for use in primary grades

AVAILABILITY OF ALTERNATE FORMS — No alternate forms are available.

ADMINISTRATION TIME	PART I: approximately 1 minute per child PART II: approximately 5 minutes per child
ADMINISTRATION PROCEDURES	Part I and Part II are administered to individual students. PART I: The examiner shows the examinee one of four pictures provided, or any other of interest. The examinee is then asked to talk about the picture in the language he or she is expected to use when learning how to read. PART II: The examiner shows the examinee one of the four pictures provided and asks him or her to tell all that comes to mind about the picture. All responses are recorded. The examiner should elicit as much language as possible from each child. No criteria are provided for determining an adequate language sample.
MATERIALS USED	PART I: examiner: teacher's manual, one picture examinee: none PART II: examiner: teacher's manual, one picture, tape recorder, record form examinee: none
SCORING PROCEDURES	PART I: The child's language sample is compared to a set of criteria produced in the manual. PART II: Taped examinee responses are transcribed onto the record form by the examiner. Transcribed responses are examined for content of specific language structures. Instructions on how to do this are provided in the manual. Frequency of use of the various structures is recorded on the form.
INTERPRETATION PROCEDURES	PART I: Based on criteria in the manual, the examiner places the examinee into one of two categories: (1) in need of oral language training, or (2) reading ready. The examiner can also

determine if Part II should be administered.

PART II: For each of the four pictures provided, there are two Language Usage Examples, one in English and one in Spanish. The teacher compares the examinee's oral language responses with a "typical" response on the same picture in the same language and evaluates the same language structure(s). This comparison yields indications of where the examinee stands with respect to a "typical" child's usage of language structures. This in turn may suggest areas of remediation.

VALIDITY — No studies are reported.

RELIABILITY — No studies are reported.

FIELD TRYOUTS — Data for the Language Usage Examples were obtained by testing approximately one hundred children. Responses were drawn from Spanish surnamed, Black, and Anglo children.

COST — Manual - $3.95

RATINGS — See pages 124-125.

THE PICTORIAL TEST OF BILINGUALISM AND LANGUAGE DOMINANCE
English/Spanish

PUBLISHER — Texas Testing Services
401 Poenisch Street
Corpus Christi, Texas 78412
(512) 991-6357; (512) 991-6810, ext. 271

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening		X	X	
Speaking	X	X	X	

DESCRIPTION	The primary objective of this test is to determine language facility in Spanish and English. Oral vocabulary is used as the basis for this determination. Results of the test are used to place children in six diagnostic categories: (1) English dominant, (2) Spanish dominant, (3) bilingual, (4) language deficient, (5) further testing advised for possible learning and language disorders or mental retardation, and (6) pseudo-bilingual. A second part to the test is provided, but its use is recommended only when additional information regarding bilingual language facility is required. Part II provides information on the child's use of language structure. It is designed for administration in English and Spanish.
AGE/GRADE LEVEL	Ages 4-8 (both parts)
AVAILABILITY OF ALTERNATE FORMS	No alternate forms are available.
ADMINISTRATION TIME	PART I: 15 minutes (estimate) PART II: not provided
ADMINISTRATION PROCEDURES	The test is administered to individual children by a bilingual examiner. PART I: The examiner shows a picture to the examinee and asks what it is. The examinee is asked to respond orally in both Spanish and English. The examiner notes responses on the answer sheet. PART II: The examiner shows the child a picture (provided in the booklet for Part II) and asks the child to tell a story of what he or she sees in English or Spanish. The child's response is recorded. The examiner then asks the child to tell the story again in the other language. This response is also recorded. A series of questions is then asked about the first picture and about a second

picture selected from the booklet. All answers are recorded.

MATERIALS USED

examiner: test manual, forty-card booklet for Part I, two-card booklet for Part II, answer sheet (If Part II, "Oral Language Production," is used, a cassette tape recorder may be needed to record responses.)

examinee: none

SCORING PROCEDURES

PART I: This section has four different scores: (1) English oral vocabulary, (2) Spanish oral vocabulary, (3) bilingual oral vocabulary, and (4) total oral vocabulary. Maximum score possible on each part is forty. Correct answers are specified in the manual. Scores are tabulated from the answer sheet.

PART II: This section designates four rating levels for both Spanish and English. The rules for classifying responses according to level, based on grammatical correctness and thought completion, are delineated in the manual.

INTERPRETATION PROCEDURES

PART I: Raw scores are converted into standard scores using the conversion tables provided. Standard scores are then used as coordinates on a diagnostic grid provided on the answer sheet. The grid classifies the student into one of six diagnostic categories mentioned in the test description. Using this information, the student can be placed in an educational program designed to meet his or her needs.

PART II: The examiner rates the student into one of four levels of oral language facility. The guidelines provided for establishing these ratings call for some subjective decisions on the part of the examiner. Part II is recommended only when additional information regarding bilingual language facility is required. It provides a check for differential ability in oral communication in English and Spanish.

VALIDITY

Content validity (Part I): The form of the test, pictures requiring oral response, was chosen based on the authors' theories on the beginnings of language development. All items used in the test were subjected to thorough analysis. Ambiguous drawings were eliminated and words shown to be more difficult in one language than in the other were rejected. The remaining item pool was then selected based on familiarity, language balance, and the relationship between grade level and knowing the correct words.

Construct validity:

1. Part I was administered to a sample group of 176 children. These children had already been classified into two groups: Spanish dominant and English dominant. Results suggest that the test does discriminate between English and Spanish dominant language preferences.

2. The two pictures that compose Part II were shown to two hundred Mexican American children (K-3). English/Spanish language responses were recorded verbatim and data were analyzed. A review of the data indicated that responses could be rated into four distinct levels of oral language facility.

Concurrent validity:

1. Because of the structure of Part I of The Pictorial Test of Bilingualism and Language Dominance (PTBLD), extremely low scores were considered potentially valuable as indicators of severe learning problems. To check this hypothesis, Part I was administered to a group of twenty-five mentally retarded Mexican American children who were chosen according to set criteria. Their mean scores were compared to a normal group of Mexican American children (n=182). t ratios were compu-

ted and indicated that total vocabulary scores differed significantly between the two groups.

2. Part I is claimed to measure the extent of oral language development in English and Spanish. To check this the Tests of General Ability (English) and the Hablidad General (Spanish) of the Inter-American Series were chosen for comparison purposes. The data indicate that English vocabulary scores on the two instruments are positively correlated. Corrections of the Spanish oral vocabulary scores indicate that different abilities were measured.

3. A random sample of twenty-five Mexican American children who had taken both the Peabody Picture Vocabulary Test and Part I of the PTBLD within a three-week time period was selected. The purpose of this investigation was to ascertain whether or not oral vocabulary scores earned by kindergarten, first, and second grade Mexican American children could be used as general indicators of average and below average tested ability. Raw score data were analyzed using product-moment correlations. There appears to be significant correlation between the English and total oral vocabulary scores on Part I of the PTBLD and the Peabody scores (significant beyond .05 level of confidence).

4. Teachers were asked to rate overall achievement of all students in their classes. Part I of the PTBLD was then administered to all students. A random sample of fifty students was selected and their scores on Part I were compared to the teacher rating. The data were analyzed using Pearson

product-moment correlation. The correlation of scores was significant (.49-.81).

5. In a test to determine if Part I of the PTBLD gave valid oral vocabulary scores, bilingual kindergarten and first grade teachers were asked to rate the oral language facility of the Mexican American students in their classes. Then Part I of the PTBLD was administered to fifty students selected at random. Validity coefficients were calculated (.52 Spanish, .85 English).

RELIABILITY

Internal consistency: A random selection of twenty-five Mexican American students at the kindergarten and first grade levels was given Part I which had been divided on an odd-even basis. Coefficients were calculated (.85 English, .89 Spanish).

Test/retest: Two student samples were randomly selected from kindergarten and first grade students; these students were retested after two-month and six-month intervals (n=50 and n=50, respectively). Stability coefficients were calculated. Coefficients were high but declined as the interval between test and retest increased (range for two-month retests .69-.96, range for six-month retest .35-.82).

No studies are reported for Part II.

FIELD TRYOUTS

Testing populations for the various validity and reliability studies have been described. A total of 468 Mexican American children from preschool through the second grade comprised the norm group for Part I. Composition of the group was as follows:

children in bilingual education classes	43%
children in migrant education programs	31%
children in regular classrooms	26%
	100%

female	54%
male	46%
	100%

All children were from south Texas.

Part II of the instrument was given to two hundred Mexican American children in kindergarten, first, second, and third grades.

COST — Specimen set - $18.50

RATINGS — See pages 124-125.

SCREENING TEST FOR AUDITORY COMPREHENSION OF LANGUAGE
English/Spanish

PUBLISHER — Learning Concepts
1501 North Lamar Boulevard
Austin, Texas 78705
(512) 474-6911

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening	X	X		
Speaking				

DESCRIPTION — The Screening Test for Auditory Comprehension of Language (STACL) was developed to provide a screening instrument for the Test for Auditory Comprehension of Language (TACL). It is designed for group administration. The test assesses oral language comprehension by looking at vocabulary and grammar. If a child falls below a certain percentile score for his or her age group, further testing with the TACL is indicated. It is designed to be used as a Spanish/English screening instrument; however, norms for interpretation of Spanish scores have not been determined.

AGE/GRADE LEVEL — Ages 3 years to 6 years 11 months

AVAILABILITY OF ALTERNATE FORMS	No alternate forms available.
ADMINISTRATION TIME	Estimated 35-45 minutes to administer both the Spanish and English versions
ADMINISTRATION PROCEDURES	To administer both language versions, the examiner must be bilingual. The STACL can be administered to groups ranging in size from four to thirty depending on age. An examiner and one or two proctors are needed to administer the test. The examiner reads a verbal stimulus; the examinees then choose one of three pictures on the appropriate page in their test booklets and cross it off. The proctor(s) aid(s) children having difficulties in finding the correct page, etc.
MATERIALS USED	examiner: test manual examinee: test, pencils proctor(s): none
SCORING PROCEDURES	One point is given for each correct answer. Correct responses are tabulated and recorded on the test form.
INTERPRETATION PROCEDURES	For the English version only, total score data can be converted to a percentile score by using a table provided in the manual. Children falling below the tenth percentile for their age group are recommended for further evaluation with the TACL. Spanish scores might be used as an indicator of bilingual proficiency.
VALIDITY	Content validity: Items used in this test are taken from the TACL and have been subjected to the analyses described for that test.
RELIABILITY	Test/retest: A sample of one hundred of four hundred children originally tested on the STACL were retested one to three weeks later. The correlation coefficient obtained was .60.
FIELD TRYOUTS	The test was administered to four hundred middle-class children between the ages of three and six

years. Each age group included sixty-eight Anglo American, twenty Black American, and twelve Mexican American children.

COST — Specimen set - $3.75

RATINGS — See pages 124-125.

TEST FOR AUDITORY COMPREHENSION OF LANGUAGE
English/Spanish

PUBLISHER — Learning Concepts
2501 North Lamar Boulevard
Austin, Texas 78705
(512) 474-6911

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening	X	X		
Speaking				

DESCRIPTION — This test is designed to measure auditory comprehension of language structures in Spanish and English, thus permitting assignment of the child to a developmental level of comprehension in each language. It can also serve as a diagnostic tool. Subscores on vocabulary, morphology, and syntax can be computed and used to pinpoint areas of difficulty for the child.

AGE/GRADE LEVEL — Norms provided for ages 3 years to 6 years 11 months (English only)

AVAILABILITY OF ALTERNATE FORMS — No alternate forms are available.

ADMINISTRATION TIME — Approximately 20 minutes

ADMINISTRATION PROCEDURES — The test is administered to individual children. The examiner must be bilingual to administer both language versions. The examiner questions

the child, who responds by pointing to the correct picture in the test booklet. The examiner records all responses on a separate scoring sheet.

MATERIALS USED

examiner: test manual with pictures, scoring sheet
examinee: none

SCORING PROCEDURES

Each correct response is given one point; total raw score is the sum of all points. Correct responses are provided on the scoring sheet.

INTERPRETATION PROCEDURES

Norms for only the English language version are available in the manual.

1. Total raw score can be interpreted in three ways through the use of tables provided in the test manual:

   (a) A developmental age can be ascertained from the converted scores.

   (b) A child's scores can be compared to those of students of the same age who comprised the norming group. Means and standard deviations for each age group are provided in the test manual.

   (c) The child's raw score can be converted to a percentile rank equivalent by age group using tables provided.

2. In addition to raw scores, language categories measured by test items (item groups) are provided in the test manual. The manual suggests that analysis of performance in these areas by children with very low total scores might provide information on specific problem areas. Scoring norms for item groups are not currently available, although ages at which

75 and 90 percent of children pass each English item are provided.

VALIDITY

Content validity: Items on the test were subjected to item analysis by the Southwest Educational Development Laboratory in Austin, Texas. A second method of item evaluation was by consensus of experts. Items were selected and retained if, by agreement of the experts, they evaluated knowledge of specific grammatical structures.

Concurrent validity: It is stated in the test manual that performance on the TACL successfully discriminates between individuals who have known disorders of language comprehension and those who do not. Several studies are mentioned and references are provided in the manual.

Construct validity:

1. An assumption underlying the development of the TACL is that increasing age is paralleled by increasing comprehension in young children. To test this assumption, a group of children between the ages of 3 and 7.25 years (n=162) was tested. Means were computed for various age levels. Results showed that age level was reflected in the test scores. t ratios that were computed between age groups were found to be significantly different.

2. If the TACL measures comprehension, then improvement in comprehension by a test group should be reflected in improved scores on the TACL, according to the hypothesis advanced by the validity study. A group of language disordered children were used in the study. Retest results of a group of dysphasic children showed significant changes after a one-year interval. Statistics are not reported in the

manual; references are provided for the research article.

RELIABILITY

Test/retest: The English version of the TACL was administered to a group of fifty-one children. The group consisted of twenty eight males and twenty-three females, twenty-five of whom were Mexican American. The Spanish TACL was administered to a group of thirty-two Mexican American children, twelve males and twenty females. All were retested within a two-week interval. The reliability scores were .94 for the English version and .93 for the Spanish version.

Kuder-Richardson: In a study of middle- and lower-class Anglo and Black children, the Kuder-Richardson reliability test was applied to the TACL. Total test reliability was .77. The study was cited and referenced but not included in the test manual.

FIELD TRYOUTS

During 1965-1966 the first version of the TACL was administered to a group of forty monolingual children (ages 2.6-6.6) in San Antonio, Texas. A revised version was then administered to 159 monolingual children ages 2.10 through 7.9. The children all had intelligence quotients above eighty and were free from severe speech or hearing problems. Further revisions were made. From 1968-1972 the TACL was used by the Southwest Educational Development Laboratory as an evaluative tool and was administered to thousands of children in various ethnic groups.

COST

Kit: Spanish - $34.95
English - $34.95

RATINGS

See pages 124-125.

Commercially Available Tests:

Sociolinguistic Tests

HOME BILINGUAL USAGE ESTIMATE
English/Spanish

PUBLISHER	R. V. Skoczylas 7649 Santa Inez Court Gilroy, California 95020 (408) 842-0759
DESCRIPTION	The Home Bilingual Usage Estimate (HBUE) is designed to measure language usage in the home. It yields a single score that is used to classify the subject in one of five categories: (1) English monolingual, (2) English dominant, (3) apparent monolingual, (4) Spanish dominant, or (5) Spanish monolingual. The test also provides information about language usage of other household members and intergenerational language use in the home. A subscale to compute three-generation bilingual usage is included in the test form.
AGE/GRADE LEVEL	Not specified; estimate primary grades
AVAILABILITY OF ALTERNATE FORMS	No alternate forms are available.
ADMINISTRATION TIME	6 minutes
ADMINISTRATION PROCEDURES	The instrument is administered individually to parents. Bilingual aides can be used to conduct the interviews. The examiner asks a series of questions designed to elicit information about the subject's language usage. The examinee responds verbally.
MATERIALS USED	examiner: HBUE scoring form (all pertinent information for administering the test is on this form) examinee: none
SCORING PROCEDURES	Responses by the examinee to questions about language usage in his or her family are assigned

a numerical value from zero to two. The basis for these number assignments is clearly explained. The examiner records these numbers on the scoring form. There are four sections to the instrument, each of which yields an English and a Spanish score (three listening sections and one speaking section). These scores are summed to get total Spanish and total English scores. The difference between total Spanish and total English is the final score.

INTERPRETATION PROCEDURES

Guidelines are provided for making a judgment using the total Spanish, total English, and final scores. Cutoff points are established for determining which of the five categories the subject will be placed in (see DESCRIPTION).

In addition to classifying individuals in the above manner, the HBUE provides a scale which, using two of the four subscores on the instrument, provides a picture of intergenerational language usage.

VALIDITY

Content validity: Items were chosen to furnish as many situations of bilingual usage in the home environment as possible. In addition, items were chosen (1) to be clear and easily understandable, (2) to elicit listening and speaking experiences of the subject, (3) to elicit responses that were as subjective as possible, and (4) to yield relative amounts of English and Spanish usage.

Concurrent validity: An experienced bilingual educator who knew the children and their families was asked to rate twenty-five children on a scale of one to five that corresponded to the ratings derived from the HBUE. He had never seen the HBUE. The HBUE was then administered to the children. The Pearson product-moment correlation coefficient between the ratings was computed ($r=.95$).

RELIABILITY

Test/retest: The instrument was administered

	twice following a four-week interval between successive administrations. The Pearson product-moment procedure was used to compute the correlation coefficient (r=.97).
FIELD TRYOUTS	The tryouts reported were the validity study and the reliability study. The only information on tryout populations was provided in the validity study, which tested twenty-five subjects.
COST	Set (twenty forms) - $4.00
RATINGS	See pages 126-127.

MORENO QUICK LANGUAGE ASSESSMENT INVENTORY
English/Spanish

PUBLISHER	Moreno Educational Company 7050 Belle Glade Lane San Diego, California 92119 (714) 461-0565
DESCRIPTION	The Moreno Quick Language Assessment Inventory (MQLAI) provides information about a student's English and Spanish language background. Its primary function is as a screening instrument for placement in English as a Second Language (ESL) or bilingual education programs. The developer claims the test functions with other language groups; however, no data regarding its use with other language groups are provided.
AGE/GRADE LEVEL	Grades K-6
AVAILABILITY OF ALTERNATE FORMS	No alternate forms are available.
ADMINISTRATION TIME	Approximately 1 minute

ADMINISTRATION PROCEDURES	The MQLAI can be administered by any school personnel. It is administered to the child's parents or guardians. The examiner asks specified questions about family history and language usage. The parent responds orally.
MATERIALS USED	examiner: student data analysis sheet, administration manual examinee: none
SCORING PROCEDURES	The examiner asks eleven questions and awards zero to one point for each answer. A final question is scored as negative points. Total points are summed and yield a total score.
INTERPRETATION PROCEDURES	The total score is compared to a four-part scale. Depending on the point total, a student is classified into one of four categories: (1) mainly English speaker, (2) transitional English speaker, (3) beginning English speaker, or (4) mainly Spanish speaker. Recommendations for placement in ESL, bilingual education, and language enrichment programs are included in category descriptions.
VALIDITY	Content validity: A multiple regression coefficient was computed using language in the classroom as the dependent variable and eleven categories of sociocultural information as independent variables. The resultant coefficient was .89. Inventory questions were then developed. Concurrent validity: Teacher ratings of students' classroom language were compared to ratings derived from testing 943 southern California students. The inventory predicted 93 percent of the classroom language for all students.
RELIABILITY	No studies are reported.
FIELD TRYOUTS	The tryout was used with 943 southern California students in grades K-6. Results are reported under VALIDITY.

96

COST Kit - $10.00

RATINGS See pages 126-127.

Tests Under Development or Undergoing Field Testing

LANGUAGE ASSESSMENT BATTERY
English/Spanish

PUBLISHER

Board of Education of the City of New York
Office of Educational Evaluation
110 Livingston Street
Brooklyn, New York 11201
(212) 596-4045

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening	X	X	X	X
Speaking	X	X	X	X

DESCRIPTION

The Language Assessment Battery (LAB) is intended to measure students' ability to use English or Spanish. There are three levels. Level 1 is designed for use with grades K-2. There are three tests in Level 1: (1) "Listening and Speaking," (2) "Reading," and (3) "Writing." Level 2 is designed for use with grades 3-6 and Level 3 for grades 7-12. There are four tests in Levels 2 and 3: (1) "Reading," (2) "Writing," (3) "Listening," and (4) "Speaking." Tests 1, 2, and 3 are objective, multiple-choice tests.

AGE/GRADE LEVEL

LEVEL 1: grades K-2
LEVEL 2: grades 3-6
LEVEL 3: grades 7-12

AVAILABILITY OF ALTERNATE FORMS

No alternate forms are available.

ADMINISTRATION TIME

LEVEL 1: Not stated

LEVEL 2: Test 1--40 minutes; Test 2--25 minutes; Test 3--40 minutes; Test 4--not stated

LEVEL 3: Test 1--40 minutes; Test 2--35 minutes; Test 3--40 minutes; Test 4--not stated

ADMINISTRATION PROCEDURES

LEVEL 1: Tests 1, 2, and 3 are administered to individual students. The examiner reads questions, shows stimulus pictures, and notes student responses on the scoring sheet. The examiner must be bilingual to administer both language versions of the test. The examinee responds by pointing, speaking, or writing.

LEVEL 2: Tests 1, 2, and 3 are administered to a group. Each test is timed. The examiner reads the test instructions aloud to the examinees as they follow in their test booklets. He or she asks two sample questions, then monitors the class during the test. On Test 3, "Listening," the examiner reads words to the test group. Examinees read or listen to test questions, choose an answer in their test booklets, and record their responses on the answer sheets.

Test 4 is administered to individual students. The examiner reads instructions, sample items, and questions to the student. He also shows stimulus pictures. The examiner scores the test as it is given and notes correct or incorrect responses on the scoring form. The examinee responds verbally to all questions.

LEVEL 3: Same as Level 2.

MATERIALS USED

LEVEL 1:
examiner: examiner's manual, stimulus picture book, answer sheet
examinee: student test booklet (Test 3), pencil

LEVEL 2: Tests 1, 2, and 3
examiner: examiner's manual
examinee: test booklet, answer sheet #1 (Tests 1 and 2), answer sheet #2 (Test 3), two pencils, eraser

Test 4
examiner: examiner's manual, stimulus pictures, answer sheet
examinee: none

LEVEL 3:
Same as Level 2, but using materials for Level 3

SCORING PROCEDURES

LEVEL 1: Tests 1 and 2 are scored as they are given. Test 3 is scored from the student test booklet. Scoring is objective--items are either right or wrong.

LEVELS 2 and 3: Tests 1, 2, and 3 are multiple-choice tests. These tests are designed to be machine scored. A scoring template is provided for hand scoring. Test 4 is scored as it is given by the examiner. Items are marked correct or incorrect based on pronunciation (Part 1), appropriateness to the stimulus (Part 2), and grammatical correctness (Part 3).

INTERPRETATION PROCEDURES

The LAB series functions as a screening instrument for a special instructional program in the New York City Public Schools. All students are given the LAB English test. Only those students scoring below the twentieth percentile for their grade level are given the LAB Spanish test. Charts are provided for converting raw scores to percentile scores for each grade level tested. Corresponding raw scores above the twenty-first percentile rank are not provided for either English or Spanish, as they do not enter into the interpretation procedures. Percentile rankings for the two tests are compared. If the percentile rank on the LAB Spanish is higher than that scored on the LAB English, the student enters a special instructional program. If the LAB English percentile score is equal to or greater than the LAB Spanish, then the student may not participate in the special program.

VALIDITY

No studies are reported.

RELIABILITY

No studies are reported.

FIELD TRYOUTS

No studies are reported.

RATINGS

See pages 126-127.

MAT-SEA-CAL ORAL PROFICIENCY TESTS
English, Spanish, Cantonese, Tagalog, Ilokano, Mandarin

PUBLISHER

Seattle Public Schools
815 Fourth Avenue North
Seattle, Washington 98109
(206) 587-5050

Center for Applied Linguistics
1611 North Kent Street
Arlington, Virginia 22209
(703) 528-4312

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening	X	X	X	X
Speaking	X	X		X

DESCRIPTION

The prototype MAT-SEA-CAL Oral Proficiency Test, an English version, was intended to determine the child's ability to understand and produce the distinctive characteristics of spoken English, to express known cognitive concepts, and to handle learning tasks in English. The prototype instrument was also designed to provide placement and instructional recommendations for a child with respect to alternate programs. Comparable tests with similar objectives were then developed for Spanish, Cantonese, Tagalog, Ilokano, and Mandarin.

The test is composed of three sections: Part I, "Listening Comprehension"; Part II, "Sentence Repetition"; and Part III, "Structured Response." The test yields two independent percentage scores, one for phonology and one for structure (morphological, syntactical, and lexical). Each section can be used independently. Part III can be used as a quick overview of the structures

	the child commands, including the phonological.
AGE/GRADE LEVEL	Grades K-4
AVAILABILITY OF ALTERNATE FORMS	No alternate forms are available.
ADMINISTRATION TIME	25 to 40 minutes, depending on the language tested
ADMINISTRATION PROCEDURES	The examiner must be fluent in the language being tested. All parts of the test except for Part I-A must be administered individually. Part I-A may be given to a group. Two of the three parts of the test in each language have a taped stimulus; the stimulus for the third ("Structured Response") is supplied by the examiner. All verbal responses are taped; written responses (marking the appropriate picture) are marked in a test booklet. The examiner sets up the recording equipment, checks to see that example questions are properly understood by the examinee, shows stimuli pictures and monitors the test. The examinee listens to the master tape and responds by either speaking into a microphone or by marking on a test booklet, depending upon the test section.
MATERIALS USED	examiner: test manual, master tape, tape recorder, stimuli pictures examinee: test booklet, tape recorder, pencil
SCORING PROCEDURES	Scoring charts are provided for each of the test's three parts. Scoring for Part I is completely objective: test booklets are marked and then recorded on the scoring chart. Parts II and III are graded on lexical, grammatical, and phonological structures. The scoring has been made as objective as possible by identifying the structure sought, isolating it from every other element in that utterance, and scoring it as correct or incorrect with respect to specific and stated criteria for the native pattern involved. Scoring is done directly

on the scoring charts. Recap sheets are provided for all parts to facilitate scoring. These sheets identify the exact language construct to be scored in each item. Parts I and III have separate boxes on their scoring charts entitled "classroom style." These identify by number sentences that are typical of a classroom setting and not of general conversation. Wrong answers are circled.

INTERPRETATION PROCEDURES

The MAT-SEA-CAI test yields a percentage proficiency score for each language tested. Language dominance can be estimated by a comparison of proficiency scores in each of the students' languages. If scores are within a ten-point spread, the student is considered a balanced speaker (high balance = both scores above 70 percent; low balance = both scores below 70 percent). If there is more than a ten-point spread between scores, the student is considered dominant in the highest scoring language. Minimal proficiency in a given language is indicated by a score of 70 percent. This figure is based on scores achieved by a control group of monolingual English speakers and was subjectively corroborated by the judgment of bilingual examiners who administered the test.

The scoring charts for Parts I, II, and III have diagnostic breakdowns in eight concept areas. A phonology breakdown is derived from Parts II and III, and a "classroom style" diagnostic profile can be computed from Parts I and III. Information from all three parts is brought together on the Student Profile Sheet. A composite percentage score for structure is computed on this sheet. A diagnostic analysis chart is included which allows the examiner to plot and draw a graph which shows the child's performance in each of the eight concept areas tested. A student profile sheet for phonology is not yet available; however, the developers are

preparing one (consult the publishers for current status). Program and curriculum planning can be effected for a class by examining, in the handbook, the grammatical manifestations corresponding to the concept areas causing the most difficulty.

VALIDITY

Content validity: Eight basic communication concepts that a child must handle in order to perform in a school setting were identified. These were considered universal. It was then determined what grammatical manifestations of the particular language (English, Spanish, Mandarin, etc.) the child must handle in order to perceive or communicate these concepts. Each language when compared to English almost always uses different language manifestations to convey basic concepts. The eight concepts are distributed throughout the three sections of the test.

A comparative study of the phonological systems of the six languages was undertaken. The variety of problems that the five languages present in contrast to English required inclusion of a large total number of phonological test items on the English version of the test.

Words used on the English version of the test are all included on high frequency word lists of English. Of forty student-generated words, all but two or three appear on such lists. The vocabulary inventories of the other languages have similar range and proportion.

Construct validity: Developmental changes reflected in age differentiation and sequential patterning of development are apparent in the scores on all six tests.

Concurrent validity: Correlation between oral proficiency as measured by the MAT-

SEA-CAL English test and overall achievement as measured by the SRA Assessment Survey has been established at the .01 significance level. Sample size and characteristics are not reported. The relationship between oral proficiency (MAT-SEA-CAL English test) and reading achievement (Metropolitan Achievement Test) was explored by comparing the scores of 147 students who had taken both tests. No students who scored below 84 percent on the oral proficiency test (84 percent corresponds to the mean score of the Seattle English speaking monolingual control group at the kindergarten level) was able to score in the fifth stanine in reading on the MAT-SEA CAL (a score in the fifth stanine indicates only a possibility of minimal success in reading).

RELIABILITY

A reliability study of the English and Spanish tests was run on data collected in eastern Washington (see FIELD TRYOUTS). The Kudor-Richardson KR-20 formula was used. Reliability for the English test was reported as r=.94 and for the Spanish test r=.91.

FIELD TRYOUTS

The six tests were field tested in the Seattle area and in several districts in eastern Washington in 1975 and in three southern Idaho school districts in 1976. Some 5,000 tests were administered. The Idaho data are not reported as they are in the process of analysis. Tests in Seattle were administered to 609 children in grades K-6 in ten elementary schools. Of the 609 students, eighty-six were a control group of monolingual English speakers chosen at random from four schools. The remaining 523 students had been identified as having a home language other than English. The schools were chosen for their large concentrations of students whose home

language was other than English and because these schools were prospective sites for bilingual education programs.

The eastern Washington sample consisted of five hundred students, 209 monolingual English speakers and 291 bilingual students in grades K-6.

RATINGS

See pages 126-127.

PORTUGUESE AS A SECOND LANGUAGE (PSL)
ENGLISH AS A SECOND LANGUAGE (ESL)
ACHIEVEMENT TESTS
English/Portuguese

PUBLISHER

Curriculum Research and Development Center
University of Rhode Island
Kingston, Rhode Island 02881
(401) 792-2824, -2992; (401) 789-8965, -8975

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening	X	X		
Speaking				

DESCRIPTION

This test is designed to measure the degree to which English or Portuguese youngsters have learned to read and understand Portuguese or English as a second language. The test has two parts, Part I, "Aural Achievement"; and Part II, "Reading Achievement."

AGE/GRADE LEVEL

Grades 2-5

AVAILABILITY OF ALTERNATE FORMS

No alternate forms are available.

ADMINISTRATION TIME

PART I: 45 minutes
PART II: 30 minutes

ADMINISTRATION PROCEDURES	This test is designed for administration to groups. The examiner should be fluent in the language being tested. Part I is a videotape presentation. Children being tested are asked to watch, listen, and react to a videotaped program. Children record answers to questions on special forms. In Part II the children are asked to read a series of eight short stories, each followed by four questions. Responses are marked in test booklets. The examiner distributes and collects the test booklets.
MATERIALS USED	examiner: master videotape, television with a wide screen examinee: pencil, answer sheet (Part I), test booklet (Part II)
SCORING PROCEDURES	Both Parts I and II are scored objectively. There is only one correct answer to each question--all others are incorrect.
INTERPRETATION PROCEDURES	Interpretation procedures were not available to reviewers.
VALIDITY	Content validity: The content of the ESL Achievement Test reflects the sequential, three-level curriculum of Mary Finocchiaro as found in her book, English as a Second Language: From Theory to Practice. For the PSL Achievement Test, a PSL curriculum on three levels had to be developed. The achievement test is based on this curriculum. The eight stories in Part II of the ESL Achievement Test are graded from primer through third. Additionally, a word list was used to control vocabulary; the source of this list was "Bucks County 1,185 Common Words" from Morten Botel's How to Teach Reading. Length and structure of sentences and stories were checked for difficulty. The content of the ESL test attempts to be culturally familiar to Portuguese students. Procedures used to construct Part II of the PSL Achievement Test are not referenced.

RELIABILITY — No studies are reported.

FIELD TRYOUTS — The ESL and PSL Achievement Tests were each administered to twenty-three students in a bilingual education program at Fox Point, Rhode Island. Frequency distributions were plotted for all scores.

RATINGS — See pages 126-127.

PORTUGUESE AS A SECOND LANGUAGE (PSL)
ENGLISH AS A SECOND LANGUAGE (ESL)
APTITUDE TESTS
English/Portuguese

PUBLISHER — Curriculum Research and Development Center
University of Rhode Island
Kingston, Rhode Island 02881
(401) 792-2824, -2922; (401) 789-8965, -8975

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening		X		X
Speaking		X		X

DESCRIPTION — These tests have been constructed to measure a Portuguese or English speaking child's ability to learn English or Portuguese, respectively, as a second language. The ESL Aptitude Test has four parts: Part I, "Structure;" Part II, "Visual Discrimination;" Part III, "Aural Discrimination;" and Part IV, "Oral Production." The PSL version has three parts; "Visual Discrimination" is not included.

AGE/GRADE LEVEL — Ages 6 to 8

AVAILABILITY OF ALTERNATE FORMS — No alternate forms are available

ADMINISTRATION TIME — PART I: 25 minutes
PART II: 5 minutes

PART III: 8 minutes
PART IV: 5 minutes

ADMINISTRATION PROCEDURES

The test must be given to individual students. The examiner must be a fluent speaker of both Portuguese and English. Additionally, the examiner needs special training. A training tape has been developed as part of the program. The examiner reads all instructions, provides all stimuli, and records all responses. The examinee responds orally.

MATERIALS USED

examiner: bags of toys, test manual, flash cards, answer sheets for each test part
examinee: bags of toys

SCORING PROCEDURES

Scoring by the examiner is objective. Correct answers are described on answer sheets and the examiner can refer to them as the test progresses. Scoring is done during the test.

INTERPRETATION PROCEDURES

Interpretation procedures were not available to reviewers.

VALIDITY

Concurrent validity: The ESL Aptitude Test was administered to seventy students in bilingual programs in New Bedford, Massachusetts, and Providence, Rhode Island, in late 1970 and early 1971. It is not specified if the revised version was used.

Scores on standard English comprehension tests for fifteen students in New Bedford and six in Providence were available. The fifteen New Bedford students had taken the Metropolitan Achievement Test. Scores on this instrument ("Word Knowledge" part) were compared to ESL Aptitude Test scores. The Pearson product-moment correlation coefficient was computed (r=.12).

The six students in Providence had taken the Gates-MacGinitie Reading Test ("Vocabulary" and "Comprehension" scores) and these scores

were compared to the ESL Aptitude Test using a Kendall's rank correlation coefficient (tau). Results were: tau=.67 (ESL Aptitude-G/M Vocabulary; and tau=.76 (ESL Aptitude-G/M Comprehension).

There were no completed studies for the PSL Aptitude Test.

RELIABILITY

No studies are reported.

FIELD TRYOUTS

The ESL Aptitude Test was used with Portuguese children in their native country. Scores obtained were analyzed and revisions were made in the test. The PSL Aptitude Test has been normed on native American children. In addition, both tests are being used in a reliability and validity study in the Fox Point (Rhode Island) Bilingual Education Program.

RATINGS

See pages 126-127.

SPANISH/ENGLISH LANGUAGE PERFORMANCE SCREENING
English/Spanish

PUBLISHER

Dr. Joyce Evans
Southwest Educational Development Laboratory
Austin, Texas 78701
(512) 476-6861

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening				
Speaking		X	X	

DESCRIPTION

This test is designed to determine a child's dominant language for initial placement in a bilingual education program. There are two parts to the test, Spanish and English, which are designed to be parallel. The test samples five language activities: (1) answering questions, (2) naming objects, (3) following directions,

	(4) describing objects, and (5) describing pictures. Test results yield one of six classifications.
AGE/GRADE LEVEL	Ages not specified; illustrative items indicate 4 to 6 years
AVAILABILITY OF ALTERNATE FORMS	No alternate forms are available.
ADMINISTRATION TIME	10 minutes, including scoring
ADMINISTRATION PROCEDURES	The test is administered to individual students by bilingual examiners. All questions and answers are oral. The examiner reads questions from the test form, but may prompt.
MATERIALS USED	examiner: test manual, recording forms, pictures (4), specified objects (toys) examinee: pictures, toys
SCORING PROCEDURES	The examiner records responses on the test forms provided. A scoring code is used so that a maximum amount of information about the child can be obtained.
INTERPRETATION PROCEDURES	Part I (Spanish) and Part II (English) are each assigned a letter (a, b, c, or d) corresponding to a description of the mode of response, which may be mostly Spanish, mixed response, mostly English, or few responses. On the basis of these data, a child may be classified into one of five categories, using a scale provided for this purpose: (1) Spanish, (2) predominantly Spanish, (3) bilingual, (4) predominantly English, or (5) English. In some cases a further classification may be needed to determine the correct category. Parts I and II are then compared and rated according to the following scale: (1) more responses to Part I, (2) approximately equal responses, (3) more responses to Part II. In certain cases (i.e., few responses on either part), a sixth

category--insufficient information to make a determination--is used. Illustrative examples of scoring procedures are supplied in the manual. The manual also provides behavioral descriptions of the language categories. The categories afford a convenient reference for grouping and teaching.

VALIDITY — No studies are reported.

RELIABILITY — No studies are reported.

FIELD TRYOUTS — No studies are reported.

RATINGS — See pages 126-127.

SWCEL COMPARATIVE LANGUAGE DOMINANCE TEST
English/Spanish

PUBLISHER — Southwestern Cooperative Educational Laboratory, Inc.
229 Truman NE
Albuquerque, New Mexico 87108
(505) 268-5348

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening	X			
Speaking				

DESCRIPTION — The SWCEL Test of Comparative Language Dominance (SWCEL-CLD) was designed to allow a comparative measure of language dominance in individuals who already speak two languages. The test has forms to measure English and Spanish of the northern New Mexico/southern Colorado vernacular. It yields scores in English and Spanish for receptive vocabulary.

AGE/GRADE LEVEL — Ages 2 to adult

AVAILABILITY OF ALTERNATE FORMS	No alternate forms are available.
ADMINISTRATION TIME	Approximately 20 minutes
ADMINISTRATION PROCEDURES	The test is administered to individual students by bilingual examiners who must speak the dialect native to the region of the children being assessed. Directions are given verbatim in either English or Spanish. The examiner then gives a stimulus word and shows a panel of four pictures. One picture corresponds to the stimulus. The examinee responds either orally by giving the number of his or her choice (1-4), or by pointing.
MATERIALS USED	examiner: book of 150 stimulus cards, SWCEL-CLD test manual, test booklet examinee: none
SCORING PROCEDURES	An age scale is provided to establish the correct starting point (selected starting point--SSP) on the 150-word list. Beginning at the SSP the examiner gives the stimulus word, shows the pictures corresponding to the word, and checks on the test booklet whether the response was correct or incorrect. A "basal" of eight consecutive correct answers is established and the test continues until there are six errors recorded in eight consecutive stimulus presentations. The last word given is the ceiling word. All words below the basal are counted correct. Incorrect responses are subtracted from the number of the ceiling word to establish the raw score. This procedure is followed for both languages of the test.
INTERPRETATION PROCEDURES	Procedures described in the test manual for determining language dominance call for establishing a differential score. Total English raw score is subtracted from total Spanish raw score. If the difference is greater than ten, the examinee is dominant in the lan-

guage in which he or she scored highest.

VALIDITY	No studies are reported.
RELIABILITY	No studies are reported.
FIELD TRYOUTS	No studies are reported.
RATINGS	See pages 126-127.

Tests Used for Experimental Purposes

OTT TEST OF ORAL LANGUAGE
English/Spanish

PUBLISHER

ETS Test Collection
Educational Testing Service
Princeton, New Jersey 08540
(609) 921-9000

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening				X
Speaking		X	X	X

DESCRIPTION

This instrument is designed to assess the level of proficiency in oral language production of young children from non-English speaking backgrounds. There are two parts to the test. Part I, "Phonemic Analysis," is administered in English. It covers common phonemic differentiations recognized as problems for Spanish speakers learning English. It yields eleven subscores and a total score. Part II, "Fluency," is administered both in English and in Spanish. Sentence structure, not grammatical correctness, is reflected in the scoring. The test yields three subscores: descriptive, inferential, and imaginative, and one subscale which is not included in total score, intonation. The test is designed for machine administration.

AGE/GRADE LEVEL

Not specified

AVAILABILITY OF ALTERNATE FORMS

No alternate forms are available.

ADMINISTRATION TIME

PART I: 5 minutes
PART II: English--7 minutes; Spanish--7 minutes

ADMINISTRATION PROCEDURES

The examiner must speak Spanish. The test is machine administered and can be given to four subjects at a time. The examiner sets up a tape cassette for each examinee, explains what is

	going to occur, and starts a master tape. Children listen to the master tape and respond orally to questions and task demands. For Part II slides are shown. The pictures advance automatically on cues from the master recording.
MATERIALS USED	A test package includes test manual, scoring guide, scoring sheets, the AUDIO Portable Multi-Media System, Model 3000 (available from Columbia Scientific Industries, Austin, Texas; instructions for use of the equipment are included in the manual), master tape recordings of the test (one for each part), and a set of slides (for Part II).
SCORING PROCEDURES	PART I: A scoring guide is used. Responses are checked for correct repetition of tested phonemes. They are marked on a scale from zero to four. This part yields eleven subscores. PART II: Scoring consists of a rate (0-4) for each item. A scoring sheet is provided. Questions are classed into three levels: (1) descriptive, (2) inferential, and (3) imaginative. An intonation subscore is computed but is not used in determining total score. A scoring guide provides instructions on scoring and scored sample items.
INTERPRETATION PROCEDURES	PART I: The eleven subscores in Part I enable the teacher to pinpoint problem areas for the child. Total scores gives an overall picture of the child's ability to reproduce difficult English phonemes. PART II: The score range on each language version is zero to ninety-six. There is no information provided on interpreting these scores.
VALIDITY	No studies are reported.
RELIABILITY	No studies are reported.
FIELD TRYOUTS	No studies are reported.

RATINGS See pages 126-127.

SPANISH/ENGLISH DOMINANCE ASSESSMENT
English/Spanish

PUBLISHER

ETS Test Collection
Educational Testing Service
Princeton, New Jersey 08540
(609) 921-9000

ORAL LANGUAGE DOMAIN COVERAGE

	Vocabulary	Grammar	Semantics	Phonology
Listening				
Speaking	X	X		

DESCRIPTION

This instrument was designed for use with children who are bilingual in Spanish and English, or monolingual in one of the two languages. It was meant to classify these children as dominant in one language or the other. There are three sections to the test: (1) language experience, (2) word availability, and (3) language sample. There are no "correct" answers to test questions. The test is given in both English and Spanish.

AGE/GRADE LEVEL

Grades 1 and 2

AVAILABILITY OF ALTERNATE FORMS

No alternate forms are available.

ADMINISTRATION TIME

Approximately 10 minutes

ADMINISTRATION PROCEDURES

The test is administered individually to the children. Two people are needed to give the test, an interviewer and a recorder. Both must be bilingual in Spanish and English--the interviewer must be a native speaker of the local dialect, the recorder must be literate in both English and Spanish as he or she writes down responses in both languages and works out scores. The

Spanish version is given first. The interviewer asks questions and provides stimuli, the child responds orally, and the recorder notes responses and computes the score.

MATERIALS USED

interviewer: test, a number of large pictures (preferably from the neighborhood) showing children, animals, etc.
recorder: interview sheets, record sheets
examinee: tape recorder and microphone

SCORING PROCEDURES

The recorder's instructions give detailed procedures for scoring Parts A, B, C, and D. It was stated that to score Part E a scoring key was needed, but this was not available for this review.

INTERPRETATION PROCEDURES

Raw scores yield three interpreted scores: (1) proficiency (English/Spanish), (2) home use (English/Spanish), and (3) word availability (English/Spanish). A description of the procedures used to compute raw scores was not available to reviewers.

VALIDITY

No studies are reported.

RELIABILITY

No studies are reported.

FIELD TRYOUTS

This instrument was developed for a bilingual education program in Pueblo, Colorado.

RATINGS

See pages 126-127.

Test Evaluations

In the matrix that follows on facing pages, each of the twenty-four tests described in the previous chapter is evaluated within the four criterion areas of measurement validity, examinee appropriateness, technical excellence, and administrative usability. As in the test descriptions, instruments are grouped into three developmental categories: (1) commercially available tests, (2) tests under development or undergoing field testing, and (3) tests used for experimental purposes. Instruments have also been classified as either language skill tests or sociolinguistic tests; however, sociolinguistic tests are restricted to the commercially available category. Where tests had two or more distinct and independent parts which received different evaluations, each part was evaluated and reported separately.

The four criterion areas appear at the top of the matrix. Criteria within each area are listed beneath the major heading. Whenever possible, criteria that deal with related aspects of a test within a single major area have also been organized into subcategories (such as format, interpretation,

etc.). Individual criteria within the same general area are separated; the four major criterion areas are set off by bold lines.

Tests are fully identified at the far left of the chart (organized alphabetically within each category) and identified by acronyms at the far right to facilitate reading right-hand pages. Points earned by tests on each criterion are recorded in the appropriate boxes. Final ratings based on the total scores obtained in each of the four major areas appear on the far right side of the chart.

Evaluation Rating Charts

FOUR CRITERION AREAS →	MEAS. VALIDITY		EXAMINEE APPROPRIATENESS									
						Format						
TEST NAME	Content and Construct	Concurrent and Predictive	Justification	Item Relevance	Instructions	Layout	Illustration/ Print Quality	Auditory Presentation	Power Test	Response Mode	Recording Responses
Rating Range →	0-7	0-4	0-1	0-4	0-4	0-1	0-1	0-1	0-1		0-2
COMMERCIALLY AVAILABLE INSTRUMENTS:											
Skill Tests											
1. Basic Inventory of Natural Language	1	0	0	4	1	1	1	0	1	Or	2
2. Bilingual Syntax Measure	3	0	0	3	4	1	1	1	1	Or	2
3. Comprehension of Oral Language	3	0	0	3	4	0	1	1	1	Wr	2
4. Del Rio Language Screening Test	5	0	0	2	4	1	1	1	1	Mi	2
5. Dos Amigos Verbal Language Scales	6	0	0	1	4	1	1	0	1	Or	2
6. Gloria & David Oral Language Assessment	4	0	0	3	1	0	0	1	0	Or	2
7. James Language Dominance Test	5	1	0	4	3	1	1	1	1	Mi	2
8. Language Dominance Criteria	0	0	0	3	1	1	1	1	1	Or	2
9. Language Facility Test (Dailey)	4	0	0	3	3	1	1	1	1	Or	2
10. Oral Language Dominance Measure	3	0	0	4	2	1	1	1	1	Or	2
11. Oral Language Evaluation – Part I	2	0	0	4	2	1	1	0	1	Or	2
– Part II	2	0	0	4	2	1	1	0	1	Or	2
12. Pictorial Test of Bilingualism and Language Dominance – Part I	7	1	0	4	3	1	1	1	1	Or	2
– Part II	3	0	0	3	3	0	1	1	1	Or	2
13. Screening Test for Auditory Comprehension of Language	5	0	0	3	4	1	1	1	1	Mi	2
14. Test for Auditory Comprehension of Language	6	2	0	1	4	1	1	1	1	Po	2

NOTE: A score of 0 means either no available information or noncompliance with the criterion.

TECHNICAL EXCELLENCE				ADMINISTRATIVE USABILITY														TOTAL GRADES				
Reliability				Administration				Interpretation										Four Criterion Areas				Test Acronyms
Alternate Form	Test-Retest	Internal Consistency	Replicability	Training of Administrator	Clarity of Manual	Number of Administrators	Scoring	Training	Range	Diversity	Score Conversion	Interpretation	Validating Group	Racial/Ethnic/Sex Representation	Can Decisions Be Made?	Alternate Forms	Form Comparability	Measurement Validity	Examinee Appropriateness	Technical Excellence	Administrative Usability	
0-3	0-3	0-2	0-1	0-1	0-1	0-1	0-2	0-1	0-1	0-1	0-2	0-1	0-1	0-2	0-2	0-1	0-1					
0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	P	F	P	P	(BINL)
0	0	0	1	1	0	1	1	1	1	0	1	1	0	2	2	0	0	P	G	P	F	(BIM)
0	0	0	1	1	0	1	2	1	0	0	1	0	0	1	0	1	0	P	G	P	P	(COL)
0	1	0	1	1	1	1	2	1	1	1	2	1	0	2	1	0	0	F	G	P	G	(DRLST)
0	0	0	1	1	1	1	2	0	1	0	2	1	0	2	1	0	0	F	G	P	F	(DAVLS)
0	0	0	1	1	0	1	1	0	1	1	0	0	0	0	1	0	0	P	P	P	P	(GDOLA)
0	3	0	1	1	1	1	2	1	1	0	2	1	0	2	1	0	0	F	G	F	F	(JLDT)
0	0	0	0	1	0	1	0	1	0	1	2	1	0	0	1	0	0	P	F	P	P	(LDC)
0	1	1	1	1	1	1	1	1	1	1	2	1	0	1	1	1	1	P	G	P	F	(LFT)
0	0	0	1	1	0	1	1	1	1	0	1	1	0	0	1	0	0	P	G	P	P	(OLDM)
0	0	0	0	1	0	1	1	1	0	0	2	1	0	1	1	1	0	P	F	P	P	(OLE-Pt 1)
0	0	0	0	1	0	1	1	1	0	0	0	0	0	1	0	0	0	P	F	P	P	(OLE-Pt 2)
0	2	1	1	1	1	1	2	1	1	0	1	1	0	2	1	0	0	F	G	F	F	(PTBLD-Pt 1)
0	0	0	0	1	0	1	0	1	1	0	2	1	0	1	1	0	0	P	F	P	P	(PTBLD-Pt 2)
0	0	0	1	1	1	1	2	1	1	0	2	1	0	1	1	0	0	P	G	P	F	(STACL)
0	2	2	1	0	1	1	2	1	1	1	2	1	0	2	1	0	0	F	G	F	G	(TACL)

G good

F fair

P poor

FOUR CRITERION AREAS →	MEAS. VALIDITY		EXAMINEE APPROPRIATENESS									
						Format						
TEST NAME	Content and Construct	Concurrent and Predictive	Justification	Item Relevance	Instructions	Layout	Illustration/ Print Quality	Auditory Presentation	Power Test	Response Mode	Recording Responses
Rating Range →	0-7	0-4	0-1	0-4	0-4	0-1	0-1	0-1	0-1		0-2
COMMERCIALLY AVAILABLE INSTRUMENTS (cont.)											
Socio-Linguistic Measures											
1. Home Bilingual Usage Estimate	3	2	0	4	4	1	1	1	1	Or	2
2. Language Dominance Criteria	1	0	0	4	4	1	1	1	1	Or	2
3. Moreno Quick Language Assessment Inventory	0	1	0	2	4	1	1	1	1	Or	2
MEASURES IN FIELD TEST STAGE											
1. Language Assessment Battery	0	0	1	3	4	1	1	1	0	Mi	1
2. MAT-SEA-CAL: Oral Proficiency Test	5	0	0	4	4	1	1	1	1	Mi	2
3. Portuguese/Eng[illegible]sh Achievement	2	0	0	4	1	0	0	1	0	Wr	2
4. Portuguese/English Aptitude	0	1	0	4	4	1	1	1	1	Or	2
5. Spanish/English Language Performance Screening	1	0	0	4	2	1	1	1	1	Or	2
6. SWCEL Comparative Language Dominance	1	0	0	2	4	1	1	1	1	Mi	2
EXPERIMENTAL MEASURES											
1. Ott Test of Oral Language	0	0	0	4	4	1	1	1	0	Or	2
2. Spanish/English Dominance Assessment	1	0	0	4	4	0	0	1	1	Or	2

NOTE: A score of 0 means either no available information or noncompliance with the criterion.

TECHNICAL EXCELLENCE				ADMINISTRATIVE USABILITY												TOTAL GRADES				
Reliability				Administration				Interpretation												
Alternate Form	Test-Retest	Internal Consistency	Replicability	Training of Administrator	Clarity of Manual	Number of Administrators	Scoring	Training	Range	Diversity	Score Conversion	Interpretation	Validating Group	Racial/Ethnic/Sex Representation	Can Decisions be Made?	Alternate Forms	Form Comparability	Four Criterion Areas: Measurement Validity	Examinee Appropriateness	Technical Excellence	Administrative Usability	Test Acronyms
0-3	0-3	0-2	0-1	0-1	0-1	0-1	0-2	0-1	0-1	0-1	0-2	0-1	0-1	0-2	0-2	0-1	0-1					
0	3	0	1	1	1	1	2	1	0	0	1	1	0	0	1	0	0	F	G	F	P	(HBUE)
0	0	0	0	1	0	1	2	1	0	1	2	1	0	0	1	0	0	P	G	P	F	(LDC)
0	0	0	1	1	0	1	2	1	1	0	2	1	0	0	2	0	0	P	G	P	F	(MDLAI)
0	0	0	0	1	0	1	2	1	1	1	1	1	0	0	1	0	0	P	G	P	F	(LAB)
0	0	0	0	1	0	1	1	0	1	1	1	0	0	0	1	0	0	P	G	P	P	(M-S-COPT)
0	0	0	0	1	0	1	2	0	0	0	0	0	0	1	0	0	0	P	F	P	P	(PEA)
0	0	0	0	1	0	1	1	0	0	0	0	0	0	1	0	0	0	P	G	P	P	(PEA)
0	0	0	0	1	0	1	0	1	0	0	2	1	0	0	2	0	0	P	G	P	P	(SELPS)
0	0	0	0	1	0	1	2	1	1	0	2	1	0	0	1	0	0	P	G	P	P	(SWCEL)
0	0	0	0	1	0	1	1	0	0	1	0	0	0	0	0	0	0	P	G	P	P	(OTOL)
0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	P	F	P	P	(SEDA)

G good

F fair

P poor

Summary

The notion of simultaneously assessing students' language performance in two or more languages is a relatively new phenomenon in the field of testing. One indication of this fact is the small number of languages for which tests are available. Of the twenty-four tests, only three can be used for languages other than Spanish: the Portuguese as a Second Language (PSL) Achievement Test, the PSL Aptitude Test, and the MAT-SEA-CAL. The former two tests may be used in both Portuguese and English; the latter test may be used to assess Mandarin, Cantonese, Tagalog, and Ilokano in addition to Spanish and English. However, these three tests are in a field test stage and are not presently available for mass distribution. Nonetheless, although most of the tests that are included in this document have a Spanish and an English component, the national search for instruments that was conducted did identify test development efforts for other language groups. Such tests were not selected for evaluation in this survey either because they do not as yet have parallel English forms or because developmental efforts are in such an early stage that

evaluation at this time would not be appropriate. However, in the interest of sharing these efforts with those who might be involved in assessing competencies of similar language groups, the names and addresses of individuals involved in developing tests for some other languages appear as an appendix to this report.

Another indication of the novelty of oral language assessment is the restricted age/grade range for which most current instruments are applicable. The majority of tests evaluated in this publication are appropriate only for primary school children. Only three oral language skill tests and one sociolinguistic test have been used with secondary school students. The Language Assessment Battery and the SWCEL Comparative Language Dominance Test have been used in grades K-12; however, both are currently undergoing field testing and are not available for mass distribution at this time. The Language Facility Test has also been used in secondary schools. However, the test originally was designed to assess English proficiency only. It has been included in this report because of its wide usage in Spanish/English bilingual programs. Finally, the Home Bilingual Usage Estimate can be used to provide sociolinguistic information on students at all grade levels; however, validation of the test has been conducted only at the primary school level, so that applicability to older students is uncertain.

The overall strength of all tests reviewed, regardless of their developmental level, resided in the criterion area of examinee appropriateness. It appears that test developers have been cognizant of the need for developing instruments that are appropriate for students from a variety of cultural backgrounds and have been successful in their attempts to meet this need.

As might be expected, those tests that are now commercially available generally scored higher than the experimental and field test instru-

ments on three of the four criterion areas (measurement validity, technical excellence, and administrative usability). However, while numerical scores for commercial tests were higher than for tests in the other two developmental categories, absolute ratings were still rather low. This holds true especially in the areas of measurement validity and technical excellence, and to a lesser extent in administrative usability.

Measurement validity and technical excellence are two criterion areas of crucial importance that have tended to be neglected in test development efforts. The aspects of validity and reliability, as evaluated by the criteria under measurement validity and technical excellence, respectively, are critical in creating accurate assessment instruments. If decisions are to be made on the basis of students' test scores, one must know what specifically the test measures and how well it accomplishes its objectives (validity). Likewise, it is important to know how stable and consistent the scores are (reliability). The tests evaluated in the present report deal primarily with classifying students into language dominance categories and identifying areas of strengths and weaknesses within specific oral language content areas. Without evidence supporting the validity and reliability of such tests, classification or instructional decisions based on test results cannot be made with confidence.

Given, then, that currently available oral language tests are limited both in languages evaluated and age/grade range covered, the question remains: What action steps can administrators take in determining how to evaluate oral language usage of different linguistic groups within their districts or regions? Three possible alternatives exist:

1. Existing tests may be utilized based on a review of the congruency between the purposes of the test(s) under consideration and the goals of district testing programs. Where there is good agreement between these two factors, districts may

select one or more tests that are presently available for use in their own programs.

2. Existing tests may be adapted to meet the needs of particular districts.

3. New tests may be developed according to the characteristics of the language group to be evaluated and the specifications established by the school administration.

In some instances, districts may not be able to find tests dealing with the language group(s) they wish to assess, or may find that no tests exist that evaluate the particular ages or grade levels in question. Thus, it may not be possible to choose alternative (1) in all cases. On the other hand, alternatives (2) and (3) are time consuming, costly, and require specialized expertise that may not be readily available.

Whichever alternative is eventually selected, other data sources must be considered as necessary supplements to test scores. All test scores are fallible; furthermore, most tests generally cover only parts of the oral language assessment domain. Therefore, sources such as classroom performance, teacher judgment, qualified professional opinion, and other test scores should always be investigated before classification and instructional decisions are made.

The need for quality assessment instruments for both students and programs is evident. Over the course of completing this survey, the authors became increasingly aware that with few exceptions, test development efforts have not recognized similar ongoing development work being conducted by other individuals. Consequently, there has been much duplication instead of extension of the testing field. To facilitate the identification of effective procedures, test developers need to be more aware of other test development efforts. Only through such collaboration can the most effective instruments be produced.

REFERENCES

[1]Elton Ridge. "Bilingual/Bicultural Summary," manuscript given at National Lau Directors' Meeting, Portland, Oregon, November 1975, pp. 1-23.

[2]United States Congress. *92nd Congress, 2nd Session, Senate Committee on Labor and Public Welfare to Amend and Extend Certain Acts Relating to Elementary and Secondary Education Programs, and for Other Purposes.* Washington, D.C.: Government Printing Office, 1974, p. 43.

[3]*Lau et al* v. *Nichols et al.* 72-6520 Sup. Ct. 1 (1974).

[4]United States Congress. House of Representatives. *Hearings Before the General Subcommittee on Education and Labor. 93rd Congress Second Session on H.R. 1085, H.R. 2490, and H.R. 11464 Bills to Amend Title VII of the Elementary and Secondary Education Act of 1965 to Extend, Improve, and Expand Programs of Bilingual Education, Teacher Training, and Child Development.* Washington, D.C.: U.S. Government Printing Office, 1974, p. 41.

[5]Department of Health, Education, and Welfare, U.S. Office of Education, *Federal Programs Supporting Educational Change, Vol. III: The Process of Change, Appendix C. Innovations in Bilingual Education.* Washington, D.C.: U.S. Government Printing Office, 1975, p. I-1.

[6]United States Congress. House of Representatives, *op cit.*, p. 5.

[7]Francis P. Dinneen, *An Introduction to General Linguistics*, New York: Holt, Rinehart and Winston, Inc., 1967, pp. 6-10.

[8]United States Statutes. Vol. 40, No. 122. Title 46 Part 123 Sec. 123.02 (n), Washington, D.C.: U.S. Government Printing Office, 1975.

[9]Perry A. Zirkel, "The Why's and Ways of Testing Bilinguality Before Teaching Bilingually." *The Elementary School Journal* 76: pp. 323-330; March 1976.

[10]Randall L. Jones, "Testing Language Proficiency in the United States Government." *Testing Language Proficiency*, Arlington: Center for Applied Linguistics, 1975, p. 2.

[11]William G. Moulton, "Linguistics and Language Teaching in the United States: 1940-1960," reprinted from *Trends in European and American Linguistics*, Netherlands: Spectrum Publishers, 1963, also available from ERIC: ED 038 879, p. 84.

[12]Calvin R. Petersen and Francis A. Carter, "Some Theoretical Problems and Practical Solutions in Proficiency Test Validity." *Testing Language Proficiency*, Arlington: Center for Applied Linguistics, 1975, p. 105.

[13]Petersen and Carter, *loc. cit.*

[14]B. Fraser, unpublished proposal. Cambridge, Mass.: Language Research Foundation, 1969, as cited by Courtney B. Cazden, *Child Language and Education*, New York: Holt, Rinehart and Winston, Inc., 1972, p. 258.

[15]Dell Hymes, "Competence and Performance in Linguistic Theory." *Language Acquisition: Models and Methods*, New York: Academic Press, 1971, pp. 3-28.

[16]Andrew D. Cohen, "The Sociolinguistic Assessment of Speaking Skills in a Bilingual Education Program." *Papers on Language Testing 1967-1974*, Washington, D.C.: Teachers of English to Speakers of Other Languages, 1975, p. 174.

BIBLIOGRAPHY

Alatis, James E., ed. *Monograph Series on Languages and Linguistics.* Report on the Twenty-First Annual Round Table Meeting on Linguistics and Language Studies: Bilingualism and Language Contact: Anthropological, Linguistic, Psychological, and Sociological Aspects. No. 23. Washington, D.C.: Georgetown University Press, 1970.

Anderson, Theodore and Boyer, Mildred. *Bilingual Education in the United States, Vol. I and Vol. II.* Washington, D.C.: U.S. Government Printing Office, 1970.

Bordie, John G. "Language Tests and Linguistically Different Learners: The Sad State of the Art." *Elementary English* 47: pp. 814-828; 1970.

Burt, Marina K. and Dulay, Heidi C., ed. *New Directions in Second Language Learning, Teaching, and Bilingual Education*, selected papers from the Ninth Annual TESOL Convention, Los Angeles, California, March 4-9, 1975. Washington, D.C.: Teachers of English to Speakers of Other Languages, 1975.

Carrow, Sister Mary A. "Linguistic Functioning of Bilingual and Monolingual Children." *Journal of Speech and Hearing Disorders 22:* pp. 371-380; 1957.

Cazden, Courtney B. *Child, Language and Education,* New York: Holt, Rinehart and Winston, Inc., 1972.

Cervenka, Edward J. *Final Report on Head Start Evaluation and Research --1966-1967 to the Institute for Educational Development.* Section VI, the Measurement of Bilingualism, etc. Texas University Rep. No. IED-66-1, 31 August 1967.

Cohen, Andrew D. "The Sociolinguistic Assessment of Speaking Skills in a Bilingual Education Program," *Papers on Language Testing 1967-1975,* Palmer, Leslie and Spolsky, Bernard, ed., Washington, D.C.: Teachers of English to Speakers of Other Languages, 1975.

Dailey, John T. *Language Facility Test. Test Administrator's Manual,* The Allington Corporation, 801 N. Pitt Street, No. 701, Alexandria, Virginia 22314. 1968.

Dinneen, Francis P. *An Introduction to General Linguistics,* New York: Holt, Rinehart and Winston, Inc., 1967.

Ervin-Tripp, Susan M. *Language Acquisition and Communicative Choice,* Dil, Anwar S., ed., Stanford: Stanford University Press, 1973.

Fishman, Joshua A., Cooper, Robert L. and Ma, Roxanna. *Bilingualism in the Barrio,* Indiana University Publications, Language Science Monograph #7. The Hague, Netherlands: Moulton and Co., 1971.

Fishman, Joshua A., ed., *Readings in the Sociology of Language,* The Hague, Netherlands: Moulton and Co., 1968.

Fraser, B. Unpublished proposal, Cambridge, Mass.: Language Research Foundation, 1969, as cited by Cazden, Courtney B., *Child Language and Education,* New York: Holt, Rinehart and Winston, Inc., 1972.

Gumperz, John J. *Language in Social Groups,* Anwar, Dil S., ed., Stanford: Stanford University Press, 1971.

Harris, David P. *Testing English as a Second Language,* New York: McGraw-Hill Book Co., 1969.

Hymes, Dell. *Language Acquisition: Models and Methods,* New York: Academic Press, 1971.

Jakobovits, Leon A. "A Functional Approach to the Assessment of Language Skills." *Journal of English as a Second Language,* 5, 2: pp. 63-76, Fall 1969.

John, Vera and Horner, J. *Early Childhood Bilingual Education,* New York: Modern Language Association, 1971.

Jones, Randall L. "Testing Language Proficiency in the United States Government." *Testing Language Proficiency,* Jones, Randall L. and Spolsky, Bernard, ed., Arlington: Center for Applied Linguistics, 1975.

Keller, Gary D. *The Bilingual Review--La Revista Bilingue,* Jamaica, New York: York College, City University of New York, 1975.

Kerlinger, Fred N., ed., and Carroll, John B., assoc. ed. *Review of Research in Education,* Vol. II. Itasca: F. E. Peacock Publishers, Inc., 1974.

Kessler, Carolyn. *The Acquisition of Syntax in Bilingual Children,* Washington, D.C.: Georgetown University Press, 1971.

Labov, William. *Sociolinguistic Patterns,* Philadelphia: University of Pennsylvania Press, 1972.

Labov, William. *The Social Stratification of English in New York City,* Washington, D.C.: Center for Applied Linguistics, 1966.

Lambert, Wallace E. and MacNamara, John. "Some Cognitive Consequences of Following a First-Grade Curriculum in a Second Language." *Journal of Educational Psychology* 60:2: pp. 86-96; 1969.

Lambert, Wallace E. and Tucker, G. Richard. *Bilingual Education of Children: the St. Lambert Experiment,* Rowley: Newbury House Publishers, Inc., 1972.

Lau et al v. *Nichols et al.* 72-6520 Sup. Ct. 1 (1974).

Lemon, Nigel. "Linguistic Development and Conceptualization--A Bilingual Study." *Journal of Cross-Cultural Psychology,* 6:2: pp. 173-188; 1975.

Mishra, Shitala P. and Hurt, Maure, Jr. "The Use of Metropolitan Readiness Tests with Mexican American Children." *California Journal of Educational Research,* pp. 182-187; 1970.

Moulton, William G. *Trends in European and American Linguistics,* Netherlands: Spectrum Publishers, 1963.

Petersen, Calvin R. and Carter, Francis A. "Some Theoretical Problems and Practical Solutions in Proficiency Test Validity." *Testing Language Proficiency,* Jones, Randall L. and Spolsky, Bernard, ed., Arlington: Center for Applied Linguistics, 1975.

Ridge, Elton, *Bilingual/Bicultural Summary,* National Lau Directors' Meeting, Portland: November 1975.

Samuda, Ronald J. *Psychological Testing of American Minorities: Issues and Consequences,* New York: Dodd, Mead & Co., 1975.

Saville, Muriel and Troike, Rudolf. *A Handbook of Bilingual*

Education. Washington, D.C.: Teachers of English to Speakers of Other Languages, 1971.

Slobin, Dan Isaac. *Studies of Child Language Development,* New York: Holt, Rinehart and Winston, Inc., 1973.

Solis, Juan P., director. *Evaluation Instruments for Bilingual Education: An Annotated Bibliography,* Austin: Dissemination Center for Bilingual-Bicultural Education, 1975.

Solis, Juan P., director. *Guide to Title VII ESEA Bilingual-Bicultural Programs: 1974-75,* Austin: Dissemination Center for Bilingual-Bicultural Education, 1975.

Stemmler, Anne O. "The LCT Language Cognition Test--A Test for Educationally Disadvantaged School Beginners." *TESOL Quarterly* 1:4: pp. 35-43; 1967.

United States Congress. House of Representatives. *Hearings Before the General Subcommittee on Education and Labor. 93rd Congress, Second Session on H.R. 2490, and H.R. 11464 Bills to Amend Title VII of the Elementary and Secondary Education Act of 1965 to Extend, Improve, and Expand Programs of Bilingual Education, Teacher Training, and Child Development.* Washington, D.C.: U.S. Government Printing Office, 1975.

United States Congress. Senate Committee. *Education Amendments of 1974. 92nd Congress, 2nd Session, Report together with Supplemental and Additional Views of the Committee on Labor and Public Welfare, United States Senate, on S. 1539 to Amend and Extend Certain Acts Relating to Elementary and Secondary Education Programs, and for Other Purposes.* Washington, D.C.: U.S. Government Printing Office, 1974.

United States Department of Health, Education, and Welfare, U.S. Office of Education. *Federal Programs Supporting Educational Change, Vol. III: The Process of Change, Appendix C. Innovations in Bilingual Education.* Washington, D.C.: U.S. Government Printing Office, 1975.

United States Statutes. Vol. 40, No. 122. Title 46 Part 123 Sec. 123.02, Washington, D.C.: U.S. Government Printing Office, 1975.

Weber, George. "Uses and Abuses of Standardized Tests in the Schools." *Occasional Papers, 22: Council for Basic Education,* Washington, D.C.: 1974.

Williams, Frederick, ed. *Language and Poverty,* Chicago: Markham Publishing Co., 1970.

Zelson, Sidney N.J. *Measurement and Evaluation of Speaking Skill in a Second Language.* ED 088 284, Department of Health, Education, and Welfare. National Institute of Education, Washington, D.C.: 1972.

Zirkel, Perry A. "A Method of Determining and Depicting Language Dominance." *TESOL Quarterly* 8: pp. 7-16; 1974.

Zirkel, Perry A. "The Why's and Ways of Testing Bilinguality Before Testing Bilingually." *The Elementary School Journal* 76. pp. 323-330; 1976.

APPENDIX

The following sources identify test development efforts in other major languages.

Language	Organization and/or Contact Person
1. Chamorro	Department of Education Government of Guam Agana, Guam 96910 Attention: Sister Ellen Jean Klein Director, Bilingual Program
2. Cherokee	Cherokee Bilingual Project P. O. Box 769 Tahlequah, Oklahoma 74464
3. Crow	Bilingual Education Program School District 17-14 522 North Center Avenue Hardin, Montana 59034
4. French	Acadiana Bilingual/Bicultural Education Project P. O. Box 156 Youngsville, Louisiana 70592 Attention: Kathleen Price Director, ESEA Title VII

		Bilingual Education Program Burgess School Berlin, New Hampshire 03570
		St. Martin Parish Instructional Center III Courville Street Breaux Bridge, Louisiana 70517 Attention: Hazel Delaheussaye Director, Bilingual Program
5.	Marshallese	Trust Territory of the Pacific Islands Office of the Higher Commissioner Saipan, Mariana Islands 96950
6.	Miccosukee	Miccosukee Bilingual Project P. O. Box 440021 Miami, Florida 33144
7.	Navaho	Bernard Spolsky The Graduate School University of New Mexico Albuquerque, New Mexico 87131
		Southwest Research Associates P. O. Box 4092 Albuquerque, New Mexico 87106
8.	Papago	Kerwo Day School Box 8 Ajo, Arizona 85321
9.	Samoan	Department of Education Pago Pago, American Samoa 96799
10.	Yup'ik	Southwest Research Associates P. O. Box 4092 Albuquerque, New Mexico 87106

The Mission of the Northwest Regional Educational Laboratory is to Help Improve Educational Practices

The Northwest Regional Educational Laboratory assists education, government, community agencies, business and labor in bringing about improvement in educational programs and processes by:

- Developing and disseminating effective educational products and procedures
- Conducting research on educational problems
- Providing technical assistance in educational problem solving
- Evaluating effectiveness of educational programs and projects
- Providing training in educational planning, management and instruction
- Serving as an information resource on effective educational programs and processes

The Laboratory is a nonprofit organization with a staff of 175 full-time employees. Over 800 member institutions participate in cooperative planning, development, evaluation and dissemination activities.

The work of the Laboratory is supported through contracts and grants with the National Institute of Education, U.S. Office of Education, other federal agencies, state education agencies, colleges, universities, local school districts, professional associations, community agencies and business.

The Northwest Regional Educational Laboratory is developing products, publications and services in these vital areas of education:

- Rural Education
- Adult Education
- Improving Teaching Competencies
- Career Education
- Competency-Based Elementary & Secondary Education
- Bilingual & Bicultural Education
- International Education
- Educational Services (For local problem solving)
- Evaluation — Audit Services
- Assessment of Student Performance
- Computer-Based Instruction & Administration
- Indian Reading & Language Development

For Information Contact:

Office of Marketing and Dissemination
Northwest Regional Educational Laboratory
710 S.W. Second Avenue
Portland, Oregon 97204
(503) 248-6960